

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization International Bureau



(43) International Publication Date
7 October 2004 (07.10.2004)

PCT

(10) International Publication Number
WO 2004/085650 A1

(51) International Patent Classification⁷: **C12N 15/50.**
C07K 14/165, A61K 39/215, G01N 33/569, A61P 31/14,
11/00

(21) International Application Number:
PCT/CN2004/000246

(22) International Filing Date: 24 March 2004 (24.03.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/457,031 24 March 2003 (24.03.2003) US
60/457,730 26 March 2003 (26.03.2003) US
60/459,931 2 April 2003 (02.04.2003) US
60/460,357 3 April 2003 (03.04.2003) US
60/461,265 8 April 2003 (08.04.2003) US
60/462,805 14 April 2003 (14.04.2003) US
60/464,886 23 April 2003 (23.04.2003) US
60/465,738 25 April 2003 (25.04.2003) US
60/470,935 14 May 2003 (14.05.2003) US

(71) Applicant (for all designated States except US): **THE UNIVERSITY OF HONG KONG [CN/CN]; G18, Eliot Hall, Pokfulam Road, Hong Kong (CN).**

(72) Inventors: **CHAN, Kwokhung; 3A, Wisteria Mansion, Taikoo Shing, Quarry Bay, Hong Kong (CN); GUAN, YI; 11A, Block 4, Pokfulam Garden, 180 Pokfulam Road, Hong Kong (CN); NICHOLLS, John, Malcolm; 5B Royalton, 118 Pokfulam Road, Pokfulam, Hong Kong (CN). PEIRIS, Joseph, Srilayal, Malik; 19/F, Block 29, Baguio Villa, 550 Victoria Road, Hong Kong (CN). POON, A1**

WO 2004/085650 A1

(54) Title: A HIGH-THROUGHPUT DIAGNOSTIC ASSAY FOR THE HUMAN VIRUS CAUSING SEVERE ACUTE RESPIRATORY SYNDROME (SARS)

(57) Abstract: The present invention relates to a high-throughput diagnostic assay for the virus causing Severe Acute Respiratory Syndrome (SARS) in humans ("hSARS virus"). In particular, the invention relates to a high-throughput reverse transcription-PCR diagnostic test for SARS associated coronavirus (SARS CoV). The present assay is a rapid, reliable assay which can be used for diagnosis and monitoring the spread of SARS and is based on the nucleotide sequences of the N (nucleocapsid)-gene of the hSARS virus. The present method eliminates false negative results and provides increased sensitivity for the assay. The invention also discloses the S (spike)-gene of the hSARS virus. The invention further relates to the deduced amino acid sequences of the N-gene and S-gene products of the hSARS virus and to the use of the N-gene and S-gene products in diagnostic methods. The invention further encompasses diagnostic assays and kits comprising antibodies generated against the N-gene or S-gene product.

Litman, Flat 21H, Block 2, PHase 3, Belvedere Garden, Tsuen Wan, Hong Kong (CN). YUEN, Kwokyung; Flat C, 19/F., Block 20, Baguio Villa, 555 Victoria Road., Hong Kong (CN). LEUNG, Frederick, C.; 24th Floor, Starr Hall, 91B Pokfulam Road, Hong Kong (CN).

(74) Agent: CHINA PATENT AGENT (H.K) LTD.; 22/F, Great Eagle Centre, 23 Harbour Road, Wanchai, Hong Kong (CN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

**A HIGH-THROUGHPUT DIAGNOSTIC ASSAY FOR
THE HUMAN VIRUS CAUSING
SEVERE ACUTE RESPIRATORY SYNDROME
(SARS)**

This application claims priority benefit to U.S. provisional application no. 60/457,031, filed March 24, 2003; U.S. provisional application no. 60/457,730, filed March 5, 26, 2003; U.S. provisional application no. 60/459,931, filed April 2, 2003; U.S. provisional application no. 60/460,357, filed April 3, 2003; U.S. provisional application no. 60/461,265, filed April 8, 2003; U.S. provisional application no. 60/462,805, filed April 14, 2003; U.S. provisional application no. 60/464,886 filed April 23, 2003, U.S. provisional application no. 60/465,738, filed April 25, 2003; and U.S. provisional application no. 60/470,935, filed 10 May 14, 2003, each of which is incorporated herein by reference in its entirety.

The instant application contains a lengthy Sequence listing which is being concurrently submitted via triplicate CD-R in lieu of a printed paper copy, and is hereby incorporated by reference in its entirety. Said CD-R, recorded on March 22, 2004, are labeled "CRF", "Copy 1" and "Copy 2", respectively, and each contains only identical 1.6 15 MB file (V9661077.APP).

1. INTRODUCTION

The present invention relates to a high-throughput diagnostic assay for the virus causing Severe Acute Respiratory Syndrome (SARS) in humans ("hSARS virus"). In 20 particular, the invention relates to a high-throughput reverse transcription-PCR diagnostic test for SARS associated coronavirus (SARS-CoV). The present assay is a rapid, reliable assay which may be used for diagnosis and monitoring the spread of SARS. The present method eliminates false negative results and provides increased sensitivity for the assay. The invention further relates to nucleotide sequences and portions thereof, useful for the 25 diagnosis of SARS. The invention further relates to nucleotide sequences and portions thereof, useful for assessing genetic diversity of SARS. The nucleotide sequences of the present invention comprise the (Nucleocapsid) N-gene and the S-gene sequences of the hSARS virus. The invention relates to a diagnostic kit that comprises nucleic acid

molecules for the detection of the N-gene or S-gene of hSARS virus. The invention also relates to the deduced amino acid sequences of the N-gene and S-gene products of the hSARS virus. The invention further relates to the use of the N-gene and S-gene products in diagnostic methods. The invention further encompasses diagnostic assays and kits

- 5 comprising antibodies generated against the N-gene or S-gene product.

2. BACKGROUND OF THE INVENTION

Recently, there has been an outbreak of atypical pneumonia in Guangdong province in mainland China. Between November 2002 and March 2003, there were 792 reported cases with 31 fatalities (WHO. Severe Acute Respiratory Syndrome (SARS) *Weekly Epidemiol Rec.* 2003; 78: 86). Patients with SARS show various clinical symptoms, including fever (of 38 degrees Celsius or above for over 24 hours), malaise, chills, headache and body ache. Chest X-rays show changes compatible with pneumonia. Other symptoms include coughing, shortness of breath or difficulty in breathing. By 3rd May 2003, a cumulative total number of 1621 cases and 179 deaths had been occurred in Hong Kong, which contributed to 26% and 41% of the global reported cases (6234) and deaths (435) respectively. As the disease is highly contagious and spreads in daily-life activities, it is important to develop a rapid and reliable diagnosis test to monitor and control the disease. In response to this crisis, the Hospital Authority in Hong Kong has increased the surveillance on patients with severe atypical pneumonia. In the course of this investigation, a number of clusters of health care workers with the disease were identified. In addition, there were clusters of pneumonia incidents among persons in close contact with those infected. The disease was unusual in its severity and its progression in spite of the antibiotic treatment typical for the bacterial pathogens that are known to be commonly associated with atypical pneumonia. The present inventors were one of the groups involved in the investigation of these patients. All tests for identifying commonly recognized viruses and bacteria were negative in these patients. Furthermore, diagnostic tests for the detection of other genes in the hSARS virus, such as the 1b-gene are not useful to accurately diagnose SARS. The disease was given the acronym Severe Acute Respiratory Syndrome ("SARS"). This virus mutates and changes rapidly and hence the diagnostic of SARS was extremely difficult until the isolation of particular regions of the virus, the N-gene and S-gene, of the hSARS virus from the SARS patients by the present inventors as disclosed herein. Namely,

the present invention discloses a diagnostic assay using particular regions in the genome of the virus for rapid, accurate, reliable and specific identification of the hSARS virus. The invention is useful in both clinical and scientific research applications. Furthermore, the present invention provides a high-throughput assay which can be used as a sensitive method
5 for diagnosis and monitoring the spread of the SARS.

3. SUMMARY OF INVENTION

The present invention is based upon the inventors' identification of a specific region of the hSARS virus, specifically, the 3' region of the hSARS viral genome, and in particular,
10 the (nucleocapsid) N-gene of the hSARS virus that may be used in diagnostic assay to detect hSARS. In particular, the N-gene is useful for the diagnosis of SARS because the N-gene has the most abundant copy number during viral infection compared to any other gene in the hSARS virus, especially when the cells are lysed. Thus, the nucleic acid sequences of the N-gene of the hSARS virus are particularly useful in a rapid and reliable diagnostic
15 assay for the hSARS virus. Furthermore, the present method eliminates false negative results and increases the sensitivity of the assay.

The hSARS virus was isolated from patients suffering from SARS in the recent outbreak of severe atypical pneumonia in China. The isolated virus is an enveloped, single-stranded RNA virus of positive polarity which belongs to the order, *Nidovirales*, of the
20 family, *Coronaviridae*. The hSARS virus is a very large RNA virus consisting of approximately 29,742 nucleotides. The complete genomic sequence of the hSARS virus was deposited in Genbank, NCBI with Accession No: AY278491 (SEQ ID NO:15), which is incorporated by reference. The isolated hSARS virus was deposited with China Center for Type Culture Collection (CCTCC) on April 2, 2003 and accorded an accession number,
25 CCTCC-V200303, as described in Section 7, *infra*, which is incorporated by reference. Also, the entire genome sequence of the hSARS virus, CCTCC-V200303, and characterization thereof are disclosed in a United States Patent Application with Attorney Docket No. V9661.0069 filed concurrently herewith on March 24, 2004, which is incorporated by reference in its entirety. The virus mutates and changes rapidly and hence
30 making the diagnostic of SARS very difficult. The present inventors have designed a diagnostic assay for detecting the presence of N-gene nucleic acid sequence or protein to rapidly, accurately, and specifically identify the hSARS virus. Furthermore, the present

inventors have designed a diagnostic assay for detecting the presence of S-gene nucleic acid sequence or protein to determine the genetic diversity of the hSARS virus. Accordingly, the invention relates to methods of detecting nucleotide sequences of the N-gene and S-gene of the hSARS virus.

5 The present invention provides a rapid, reliable assay for the detection of hSARS virus. In preferred embodiment, the detection of hSARS virus includes the use of the nucleic acid molecules of the present invention in a polymerase chain reaction, Reverse transcription-Polymerase Chain Reaction (RT-PCR), Southern analysis, Northern analysis, or other nucleic acid hybridization for the detection of hSARS nucleic acids. In one
10 embodiment, the invention provides methods for detecting the presence, activity or expression of the hSARS virus of the invention in a biological material, such as cells, nasopharyngeal aspirate, sputum, blood, saliva, urine, stool and so forth. In preferred embodiments, the biological material is nasopharyngeal aspirate or stool. The increased or decreased activity or expression of the hSARS virus in a sample relative to a control sample
15 can be determined by contacting the biological material with an agent which can detect directly or indirectly the presence, activity or expression of the hSARS virus. In a specific embodiment, the detecting agents are the nucleic acid molecules of the present invention.

The present invention also relates to a method for identifying a subject infected with the hSARS virus, said method comprising obtaining total RNA from a biological sample
20 obtained from the subject; reverse transcribing the total RNA to obtain cDNA; and subjecting the cDNA to PCR assay using a set of primers derived from a nucleotide sequence of the hSARS. In preferred embodiments, the primers are derived from the (nucleocapid) N-gene. In most preferred embodiments, the primers comprise the nucleotide sequences of SEQ ID NOS:2475 and/or 2476 or SEQ ID NOS:2480 and/or 2481. In
25 another preferred embodiments, the primers are derived from the (spike) S-gene. In more preferred embodiments, the primers comprise the nucleotide sequences of SEQ ID NOS:2477 and/or 2478.

The invention further relates to the use of the sequence information of the isolated virus for diagnostic and therapeutic methods. In a specific embodiment, the invention
30 provides nucleic acid molecules which are suitable for use as primers consisting of or comprising the nucleotide sequence of SEQ ID NO:1, 11, 13, 15, 2471, or 2473, or a complement thereof, or at least a portion of the nucleotide sequence thereof. In the most preferred embodiment, the primers comprise the nucleic acid sequence of SEQ ID

NOS:2475 and/or 2476 or SEQ ID NOS:2480 and/or 2481 for the detection of N-gene. In another most preferred embodiment, the primers comprise the nucleic acid sequence of SEQ ID NO:2477 and/or 2478 for the detection of S-gene. In another specific embodiment, the invention provides nucleic acid molecules which are suitable for hybridization to hSARS

5 nucleic acid, including, but not limited to, as PCR primers, Reverse Transcriptase primers, probes for Southern analysis or other nucleic acid hybridization analysis for the detection of hSARS nucleic acids, *e.g.*, consisting of or comprising the nucleotide sequence of SEQ ID NO:1, 11, 13, 15, 2471, 2473, 2475, 2476, 2477, 2478, 2480 or 2481, or a complement thereof, or a portion thereof. In a preferred embodiment, primers that amplify fragments

10 comprising (nucleotide position 18057 to 18222 or portions thereof of SEQ ID NO:15) 1b gene; (nucleotide position 21920-22107, or portions thereof of SEQ ID NO:15) M-gene; and (nucleotide position 28658-28883, or portions thereof of SEQ ID NO:15) N-gene may be used for probe synthesis for detection of hSARS nucleic acids. In a specific embodiment, the invention provides a diagnostic kit comprising nucleic acid molecules which are suitable

15 for use to detect the N-gene of hSARS. In a specific embodiment, the N-gene comprises nucleic acid sequence of SEQ ID NO: 2471. In specific embodiment, the nucleic acid molecules comprise nucleic acid sequence of SEQ ID NOS:2475 and/or 2476 or SEQ ID NOS:2480 and/or 2481. In preferred embodiments, the diagnostic kit further comprises a control for amplification of 1b gene. In specific embodiments, the primers used for

20 amplifying 1b gene are SEQ ID NOS:3 and/or 4. In another specific embodiments, the diagnostic kit further comprises an internal control using pig β -actin gene. In specific embodiments, the primers used for amplifying β -actin gene are SEQ ID NOS:2482 and/or 2483.

In a specific embodiment, the invention provides a diagnostic kit comprising nucleic acid molecules which are suitable for use to detect the S-gene of hSARS. In a specific embodiment, the S-gene comprises nucleic acid sequence of SEQ ID NO: 2473. In specific embodiment, the nucleic acid molecules comprise nucleic acid sequence of SEQ ID NOS: 2477 and/or 2478. The invention further encompasses chimeric or recombinant viruses encoded in whole or in part by said nucleotide sequences.

30 In another specific embodiment, the invention provides nucleic acid molecules comprising a nucleotide sequence of SEQ ID NO:1, 11, 13, 2471, and/or 2473. In a specific embodiment, the present invention provides isolated nucleic acid molecules comprising or, alternatively, consisting of the nucleotide sequence of SEQ ID NO:1, a complement thereof

or a portion thereof, preferably at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:1, or a complement thereof. In another specific embodiment, the present invention provides isolated nucleic acid molecules comprising or, alternatively, consisting of the nucleotide sequence of SEQ ID NO:11, a complement thereof or a portion thereof, preferably at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200, or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:11, or a complement thereof. In yet another specific embodiment, the present invention provides isolated nucleic acid molecules comprising or, alternatively, consisting of the nucleotide sequence of SEQ ID NO:13, a complement thereof or a portion thereof, preferably at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:13, or a complement thereof. In another specific embodiment, the present invention provides isolated nucleic acid molecules comprising or, alternatively, consisting of the nucleotide sequence of SEQ ID NO:15, a complement thereof or a portion thereof, preferably at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, 21,000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000 or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:15, or a complement thereof. In another specific embodiment, the present invention provides isolated nucleic acid molecules comprising or, alternatively, consisting of the nucleotide sequence of SEQ ID NO:2471, a complement thereof or a portion thereof, preferably at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200 or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2471, or a complement thereof. In another specific embodiment, the present invention provides isolated nucleic acid molecules comprising or, alternatively, consisting of the nucleotide sequence of SEQ ID NO:2473, a complement thereof or a portion thereof, preferably at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200, 2,000, 3,000, or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2473, or a complement thereof. Furthermore, in another specific

embodiment, the invention provides isolated nucleic acid molecules which hybridize under stringent conditions, as defined herein, to a nucleic acid molecule having the sequence of SEQ ID NO:1, 11, 13, 15, 2471, or 2473, or a complement thereof. In one embodiment, the invention provides an isolated nucleic acid molecule which is antisense to the coding strand
5 of a nucleic acid of the invention. In another specific embodiment, the invention provides isolated polypeptides or proteins that are encoded by a nucleic acid molecule comprising or, alternatively consisting of a nucleotide sequence that is at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:1, or a complement thereof. In yet another specific
10 embodiment, the invention provides isolated polypeptides or proteins that are encoded by a nucleic acid molecule comprising or, alternatively consisting of a nucleotide sequence that is at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200 or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:11, or a complement thereof. In yet
15 another specific embodiment, the invention provides isolated polypeptides or proteins that are encoded by a nucleic acid molecule comprising or, alternatively consisting of a nucleotide sequence that is at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:13, or a complement thereof. In yet another specific embodiment,
20 the invention provides isolated polypeptides or proteins that are encoded by a nucleic acid molecule comprising or, alternatively consisting of a nucleotide sequence that is at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000,
25 19,000, 20,000, 21,000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000 or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:15, or a complement thereof. In yet another specific embodiment, the invention provides isolated polypeptides or proteins that are encoded by a nucleic acid molecule comprising or, alternatively consisting of a nucleotide sequence that is at least 5, 10, 15, 20, 25, 30, 35, 40,
30 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200 or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2471, or a complement thereof. In yet another specific embodiment, the invention provides isolated polypeptides or proteins that are encoded by a nucleic acid

- molecule comprising or, alternatively consisting of a nucleotide sequence that is at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200, 2,000, 3,000, or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2473, or a complement thereof. The invention further provides proteins or polypeptides that are isolated from the hSARS virus, including viral proteins isolated from cells infected with the virus but not present in comparable uninfected cells. The invention further provides proteins or polypeptides shown in Figures 11 (SEQ ID NOS:17-239, 241-736 and 738-1107) and 12 (SEQ ID NOS:1109-1589, 1591-1964 and 1966-2470). The invention further provides proteins or polypeptides having the amino acid sequence of SEQ ID NO:2472 or 2474. The polypeptides or the proteins of the present invention preferably have a biological activity of the protein (including antigenicity and/or immunogenicity) encoded by the sequence of SEQ ID NO:1, 11, 13, 2471, or 2473. In other embodiments, the polypeptides or the proteins of the present invention have a biological activity of the protein (including antigenicity and/or immunogenicity) encoded by a nucleotide sequence that is at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, 21,000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000 or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:15, or a complement thereof. In other embodiments, the polypeptides or the proteins of the present invention have a biological activity of the protein (including antigenicity and/or immunogenicity) of Figures 11 (SEQ ID NOS:17-239, 241-736 and 738-1107) and 12 (SEQ ID NOS:1109-1589, 1591-1964 and 1966-2470). The invention further provides proteins or polypeptides having a biological activity of the protein having amino acid sequence of SEQ ID NO: 2472 or 2474.

In one aspect, the invention provides a method for propagating the hSARS virus in host cells comprising infecting the host cells with the isolated hSARS virus, culturing the host cells to allow the virus to multiply, and harvesting the resulting virions. Also provided by the present invention are host cells that are infected with the hSARS virus.

In one aspect, the invention relates to the use of the isolated hSARS virus for diagnostic and therapeutic methods. In a specific embodiment, the invention provides a method of detecting in a biological sample an antibody immunospecific for the hSARS

virus using the isolated hSARS virus or any proteins or polypeptides thereof. In another specific embodiment, the invention provides a method of screening for an antibody which immunospecifically binds and neutralizes hSARS. Such an antibody is useful for a passive immunization or immunotherapy of a subject infected with hSARS.

- 5 The invention further provides antibodies that specifically bind a polypeptide of the invention encoded by the nucleotide sequence of SEQ ID NO:1, 11, 13, 2471, or 2473, or a fragment thereof, or encoded by a nucleic acid comprising a nucleotide sequence that hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO:1, 11, 13, 2471, or 2473, and/or any hSARS epitope, having one or more biological activities of a 10 polypeptide of the invention. The invention further provides antibodies that specifically bind polypeptides of the invention encoded by the nucleotide sequence of SEQ ID NO:15, or a fragment thereof. These polypeptides include those shown in Figures 11 (SEQ ID NOS:17-239, 241-736 and 738-1107) and 12 (SEQ ID NOS:1109-1589, 1591-1964 and 1966-2470). In another embodiment, the polypeptide comprises amino acid sequence of 15 SEQ ID NO:2472, or 2474. The invention further provides antibodies that specifically bind polypeptides of the invention encoded by a nucleic acid comprising a nucleotide sequence that hybridizes under stringent conditions to the nucleotide sequence of SEQ ID NO:15, and/or any hSARS epitope, having one or more biological activities of a polypeptide of the invention. Such antibodies include, but are not limited to polyclonal, monoclonal, bi- 20 specific, multi-specific, human, humanized, chimeric antibodies, single chain antibodies, Fab fragments, F(ab')_n fragments, disulfide-linked Fvs, intrabodies and fragments containing either a VL or VH domain or even a complementary determining region (CDR) that specifically binds to a polypeptide of the invention.

In another embodiment, the invention provides vaccine preparations, comprising the 25 hSARS virus, including recombinant and chimeric forms of said virus, or protein subunits of the virus. In a specific embodiment, the vaccine preparations of the present invention comprise live but attenuated hSARS virus with or without adjuvants. In another specific embodiment, the vaccine preparations of the invention comprise an inactivated or killed hSARS virus. Such attenuated or inactivated viruses may be prepared by a series of 30 passages of the virus through the host cells or by preparing recombinant or chimeric forms of virus. Accordingly, the present invention further provides methods of preparing recombinant or chimeric forms of hSARS. In another specific invention, the vaccine preparations of the present invention comprise a nucleic acid or fragment of the hSARS

virus, *e.g.*, the virus having accession no. CCTCC-V200303, or nucleic acid molecules having the sequence of SEQ ID NO. 1, 11, 13, 15, 2471 or 2473, or a fragment thereof. In another embodiment, the invention provides vaccine preparations comprising one or more polypeptides isolated from or produced from nucleic acid of hSARS virus, for example, of deposit accession no. CCTCC-V200303. In a specific embodiment, the vaccine preparations comprise a polypeptide of the invention encoded by the nucleotide sequence of SEQ ID NO:1, 11, 13, 2471 or 2473, or a fragment thereof. In a specific embodiment, the vaccine preparations comprise polypeptides of the invention as shown in Figures 11 (SEQ ID NOS:17-239, 241-736 and 738-1107) and 12 (SEQ ID NO:1109-1589, 1591-1964 AND 1966-2470) or encoded by the nucleotide sequence of SEQ ID NO:15, or a fragment thereof. In a specific embodiment, the vaccine preparations comprise polypeptides comprising amino acid sequence of SEQ ID NO:2472 or 2474. Furthermore, the present invention provides methods for treating, ameliorating, managing or preventing SARS by administering the vaccine preparations or antibodies of the present invention alone or in combination with adjuvants, or other pharmaceutically acceptable excipients.

In another aspect, the present invention provides pharmaceutical compositions comprising anti-viral agents of the present invention and a pharmaceutically acceptable carrier. In a specific embodiment, the anti-viral agent of the invention is an antibody that immunospecifically binds hSARS virus or any hSARS epitope. In preferred embodiments, such antibodies neutralize the hSARS virus. In a specific embodiment, the anti-viral agent of the invention binds a fragment, variant, homolog of N-gene or S-gene of hSARS virus. In a specific embodiment, the anti-viral agent of the invention binds a fragment, variant, homolog of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2472 or 2474. In another specific embodiment, the anti-viral agent is a polypeptide or protein of the present invention or nucleic acid molecule of the invention. The invention also provides kits containing a pharmaceutical composition of the present invention.

3.1 Definitions

The term “an antibody or an antibody fragment that immunospecifically binds a polypeptide of the invention” as used herein refers to an antibody or a fragment thereof that immunospecifically binds to the polypeptide encoded by the nucleotide sequence of SEQ ID NO:1, 11, 13, 15, 2471 2473, or the polypeptides shown in Figures 11 and 12, or a fragment

thereof, and does not non-specifically bind to other polypeptides. An antibody or a fragment thereof that immunospecifically binds to the polypeptide of the invention may cross-react with other antigens. Preferably, an antibody or a fragment thereof that immunospecifically binds to a polypeptide of the invention does not cross-react with other

- 5 antigens. An antibody or a fragment thereof that immunospecifically binds to the polypeptide of the invention, can be identified by, for example, immunoassays or other techniques known to those skilled in the art.

An "isolated" or "purified" peptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is

- 10 derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of a polypeptide/protein in which the polypeptide/protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, a polypeptide/protein that is substantially free of cellular material includes preparations of the
15 polypeptide/protein having less than about 30%, 20%, 10%, 5%, 2.5%, or 1%, (by dry weight) of contaminating protein. When the polypeptide/protein is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When
20 polypeptide/protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly, such preparations of the polypeptide/protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than polypeptide/protein fragment of interest. In a preferred embodiment of the present invention, polypeptides/proteins are
25 isolated or purified.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule.

Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by
30 recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In a preferred embodiment of the invention, nucleic acid molecules encoding polypeptides/proteins of the invention are isolated or purified. The term "isolated" nucleic acid molecule does not include a nucleic acid that is a member of a

library that has not been purified away from other library clones containing other nucleic acid molecules.

- The term "portion" or "fragment" as used herein refers to a fragment of a nucleic acid molecule containing at least about 25, 30, 35, 40, 45, 100, 150, 200, 250, 300, 350, 400,
- 5 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, 21,000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000, or more contiguous nucleic acids in length of the relevant nucleic acid molecule and having at least one functional feature of the
- 10 nucleic acid molecule (or the encoded protein has one functional feature of the protein encoded by the nucleic acid molecule); or a fragment of a protein or a polypeptide containing at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 400, 500, 600, 700, 800, 900, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 4,100, 4,200, 4,300, 4,350, 4,360, 4,370,
- 15 4,380 amino acid residues in length of the relevant protein or polypeptide and having at least one functional feature of the protein or polypeptide.

The term "3' region of the hSAR viral genome" refers to from about nucleotide position 18,000 to 29742 of SEQ ID NO:15.

- The term "having a biological activity of the protein" or "having biological activities of the polypeptides of the invention" refers to the characteristics of the polypeptides or proteins having a common biological activity similar or identical structural domain and/or having sufficient amino acid identity to the polypeptide encoded by the nucleotide sequence of SEQ ID NO:1, 11, 13, 15, 16, 240, 737, 1108, 1590, 1965, 2471 or 2473. Such common biological activities of the polypeptides of the invention include antigenicity and immunogenicity.

- The term "under stringent condition" refers to hybridization and washing conditions under which nucleotide sequences having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identity to each other remain hybridized to each other. Such hybridization conditions are described in, for example but not limited to, Current
- 30 Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.; Basic Methods in Molecular Biology, Elsevier Science Publishing Co., Inc., N.Y. (1986), pp. 75-78, and 84-87; and Molecular Cloning, Cold Spring Harbor Laboratory, N.Y. (1982), pp. 387-389, and are well known to those skilled in the art. A preferred, non-limiting example

of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC), 0.5% SDS at about 68°C followed by one or more washes in 2X SSC, 0.5% SDS at room temperature. Another preferred, non-limiting example of stringent hybridization conditions is hybridization in 6X SSC at about 45°C followed by one or more washes in 5 0.2X SSC, 0.1% SDS at about 50-65°C.

The term "variant" as used herein refers either to a naturally occurring genetic mutant of hSARS or a recombinantly prepared variation of hSARS each of which contain one or more mutations in its genome compared to the hSARS of CCTCC-V200303. The term "variant" may also refer either to a naturally occurring variation of a given peptide or 10 a recombinantly prepared variation of a given peptide or protein in which one or more amino acid residues have been modified by amino acid substitution, addition, or deletion.

4. BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows a partial DNA sequence (SEQ ID NO:1) and its deduced amino acid 15 sequence (SEQ ID NO:2) obtained from the SARS virus that has 57% homology to the RNA-dependent RNA polymerase protein of known *Coronaviruses*.

Figure 2 shows an electron micrograph of the novel hSARS virus that has similar morphological characteristics of coronaviruses.

Figure 3 shows an immunofluorescent staining for IgG antibodies that are 20 specifically bound to the FrHK-4 cells infected with the novel human respiratory virus of *Coronaviridae*.

Figure 4 shows an electron micrograph of ultra-centrifuged deposit of hSARS virus that was grown in the cell culture and negatively stained with 3% potassium phosphotungstate at pH 7.0.

25 Figure 5A shows a thin-section electron micrograph of lung biopsy of a patient with SARS; and Figure 5B shows a thin section electron micrograph of hSARS-infected cells.

Figure 6 shows the result of phylogenetic analysis for the partial protein sequence (215 amino acids; SEQ ID NO:2) of the hSARS virus (GenBank accession number AY268070). The phylogenetic tree is constructed by the neighbor-jointing method. The 30 horizontal-line distance represents the number of sites at which the two sequences compared are different. Bootstrap values are deducted from 500 replicates.

Figure 7A shows an amplification plot of fluorescence intensity against the PCR cycle in a real-time quantitative PCR assay that can detect a hSARS virus in samples quantitatively. The copy numbers of input plasmid DNA in the reactions are indicated. The X-axis denotes the cycle number of a quantitative PCR assay and the Y-axis denotes the fluorescence intensity (F1) over the background. Figure 7B shows the result of a melting curve analysis of PCR products from clinical samples. Signals from positive (+ve) samples, negative (-ve) samples and water control (water) are indicated. The X-axis denotes the temperature (°C) and the Y-axis denotes the fluorescence intensity (F1) over the background.

Figure 8 shows another partial DNA sequence (SEQ ID NO:11) and its deduced amino acid sequence (SEQ ID NO:12) obtained from the SARS virus.

Figure 9 shows yet another partial DNA sequence (SEQ ID NO:13) and its deduced amino acid sequence (SEQ ID NO:14) obtained from the SARS virus.

Figure 10 shows the entire genomic DNA sequence (SEQ ID NO:15) of the SARS virus.

Figure 11 shows the deduced amino acid sequences obtained from SEQ ID NO:15 in three frames (*see* SEQ ID NOS:16, 240 and 737). An asterisk (*) indicates a stop codon which marks the end of a peptide. The first-frame amino acid sequences: SEQ ID NOS:17-239; the second-frame amino acid sequences: SEQ ID NOS:241-736; and the third-frame amino acid sequences: SEQ ID NO:738-1107.

Figure 12 shows the deduced amino acid sequences obtained from the complement of SEQ ID NO:15 in three frames (*see* SEQ ID NOS:1108, 1590 and 1965). An asterisk (*) indicates a stop codon which marks the end of a peptide. The first-frame amino acid sequences: SEQ ID NOS:1109-1589; the second-frame amino acid sequences: SEQ ID NOS:1591-1964; and the third-frame amino acid sequences: SEQ ID NO:1966-2470.

Figure 13 shows the N-gene primer sequences (which amplifies nucleotides at position 29247-29410 of SEQ ID NO:2471). 150# (SEQ ID NO:2475); 200# (SEQ ID NO:2476); and S-gene primer sequences (which amplifies nucleotides at position 24751 to 25049 of SEQ ID NO:2473). 131# (SEQ ID NO:2477); 132# (SEQ ID NO:2478).

Figure 14A shows the nucleic acid sequence of the N-gene (SEQ ID NO:2471). Figure 14B shows the amino acid sequence of the N-gene (SEQ ID NO:2472).

Figure 15A shows the nucleic acid sequence of the S-gene (SEQ ID NO:2473). Figure 15B shows the amino acid sequence of the S-gene (SEQ ID NO:2474).

Figure 16 shows the genome organization and transcription strategy of SARS-CoV HK-39. Genomic and mRNA transcripts are capped (black circles), carry leader sequences (vertical lines) at 5' proximal and are polyadenylated (A¹⁵). Arrows point the position of the intergenic sequence, 5'-CTAAACGAAC-3' (SEQ ID NO:2479). After release of the

- 5 positive-sense genomic RNA in the cytoplasm of host cell, the viral RNA-dependent RNA polymerase, encoded from ORF 1a and 1b, is synthesized. It carries out transcription of a full-length complementary (negative-sense) RNA, from which new genomic RNA, an overlapping set of subgenomic mRNA transcripts, and leader RNA are synthesized. Note that all transcripts are preceded with common 5' leader sequences and common 3' ends.
- 10 ORF1a and 1b - RNA-dependent RNA polymerase; S - the major peplomer glycoprotein; M - transmembrane glycoprotein; N - nucleocapsid; X1, X2, X3 - putative proteins.

Figure 17 shows a construct map of pSARSCoV-ORF1b-N. PCR products amplified from ORF1b (1b) and N gene of SARS-CoV were co-ligated into a cloning vector pCR2.1-TOPO (Invitrogen). The nucleotide (nt) numbers corresponds to the positions in
15 the sequence of HK-39 strain SARS-CoV (AY278491). Shadowed areas indicate the amplicons by the primers used in diagnostic test (*i.e.*, SEQ ID NOS:2480 and 2481).

Figures 18 shows a photo of an agarose gel after electrophoresis of total RNA extracted from SARS patients using SV Total RNA isolation system. The extracted RNA was then subjected to a reverse-transcription polymerase chain reaction (RT-PCR) assay for
20 the detection of coronavirus in the patients.

Figure 19 shows the effect of potential inhibitors in Reverse Transcription Polymerase Chain Reaction (RT-PCR). To remove potential inhibitors, total RNA eluted from SV96 Binding Plate was precipitated with 95 % ethanol and 3 M sodium acetate and resuspend in 12 μ l of nuclease-free water. RT-PCR was performed with actin-F (SEQ ID
25 NO:2482) and actin-R (SEQ ID NO:2483) primers. Numbers indicated were the number of pig kidney epithelial (PK-15) cell added in the sample as an internal control. There was no DNA fragment amplified with untreated RNA samples.

Figure 20 shows the primers used for amplifying various genes. SRS251 (SEQ ID NO:2480) and SRS252 (SEQ ID NO:2481) amplified a 225 base pair fragment from the
30 region of N-gene that showed no homology to other coronavirus. coro3 (SEQ ID NO:3) and coro4 (SEQ ID NO:4) amplified RNA-dependent RNA polymerase (1b gene) as a control. Actin-F (SEQ ID NO:2482) and actin-R (SEQ ID NO:2483) amplified a 745 base pair fragment from β -actin gene as internal control for PCR assays.

Figure 21A shows Amplification plot of fluorescence Intensity against the number of PCR cycles. Black lines show the dynamic range of N-gene specific PCR with serially diluted plasmid construct from 10^1 to 10^6 copies. NPA samples from non-SARS patients, including patients suffering from adenovirus ($n = 5$), respiratory syncytial virus ($n = 5$),
5 human metapneumovirus ($n = 5$), influenza A virus ($n = 5$), or influenza B virus ($n = 5$) infection, are shown in gray lines. Lines with triangles denotes the SARS-CoV positive NPA samples; NTC represents no template control; X-axis indicates the cycle number of quantitative PCR performed, while Y-axis represents the fluorescence intensity (FAM-400) over background signal (Delta Rn). Inlet shows the melting curve analysis of the PCR
10 products. Signals from positive (+ve), negative (-ve) samples and no template control are indicated. X-axis indicates the temperature (°C), while Y-axis represents the fluorescence intensity (Delta Rn). Figure 21B shows comparison of dynamic ranges of N-gene and 1b-gene specific PCRs. Dynamic ranges of both N-gene and 1b-gene PCR were obtained with same plasmid construct in which 1:1 ratio of corresponding amplicons were subcloned.
15 Serially diluted plasmid with copy number ranged from 10^1 to 10^5 copies was used as template in both PCRs. Lines with triangles denotes N-gene specific PCR while the gray lines indicates 1b-gene specific PCR. Inlet shows Ct values \pm standard deviation in triplicate set of experiment of both PCRs with different copy numbers of template used. NTC represents no template control; X-axis indicated the cycle number of quantitative PCR
20 performed, while Y-axis represents fluorescence intensity.

Figure 22A and 22B show an amplification curve and a melting curve, respectively, of real-time quantitative PCR specific to 1b (using the primers having SEQ ID NOS:3 and 4) and N gene (using the primers having SEQ ID NOS:2480 and 2481) of SARS CoV. Fig. 22A: Amplification plot of fluorescence intensity against the number of PCR cycles. One (1) μ l
25 of cDNA from a NPA, tracheal dispersion and lung biopsy of patients with clinical symptoms were used as template in each PCR. Fifty (50) cycles of PCR were performed to achieve the saturation phase of the reaction. X-axis indicates the cycle number of quantitative PCR performed, while y-axis represents the fluorescence intensity (FAM-490) over background signal. Horizontal gray line indicates the calculated threshold value by
30 maximum curvature approach, and the baseline cycle Ct was calculated automatically. Inlet shows half-maximal fluorescence value (1/2 max) and Ct of both PCR with cDNA from various tissue isolated from a key patient (patient A indicated in *New Engl. J. Med.* 348:1967-76 (by Drosten *et al.*, 2003) in three different time points. NPA = nasopharyngeal

aspirate; TW = tracheal wash; LW = lung wash. Fig. 22B: Melting curves of PCR products. Melting curve analysis was carried out after 10-minute further-extension step of the reaction. The temperature was raised from 56°C to 94°C by 76 increments of 0.5°C each, while each set-point temperature had been held for 7 seconds for data collection and analysis. Melting
5 temperature of 1b- and N-gene specific PCR products was 80.5°C and 85.5°C respectively. X-axis indicates the temperature in degree Celsius while Y-axis represents the fluorescence intensity (FAM-490) over background signal. One (1) µl of water was used as no template control in the reaction.

Figure 23 shows the diagnostic result of 48 clinical samples using the primers
10 having SEQ ID NOS:2480 and 2481, respectively, with β-actin PCR as an internal control. Upper bands in each row showed a 745 bp DNA fragment amplified with actin-F and actin-R, while lower bands were the amplicons by primers specific for N-gene of SARS coronavirus (225 bp). -ve control (water) and +ve control (cDNA from SARS coronavirus infected vero cell) of the assay were indicated. Five (5) µl of PCR products of both
15 reactions were mixed and loaded into the sample well in a 2 % agarose gel. M = 1 kb plus molecular marker (Invitrogen).

Figure 24 shows Northern Blot analysis of SARS-CoV total RNA. Total RNA of SARS-CoV was extracted from SARS-CoV infected Vero E6 cell. RNA was separated in a 1 % denaturing gel containing 6.29 % formaldehyde. Afterwards RNA was transferred to
20 positively charged nylon membrane and hybridized with digoxigenin-labelled PCR fragments specific to 1b, S, M and N genes, respectively. Lane 1 – 1b; lane 2 – S; lane 3 – M; lane 4 – N. Vertical bar showed the molecular size reference. Arrows indicates the transcripts hybridized with N probe. Signals were analyzed by chemiluminescence.

Figure 25 shows the DNA probes used in Nothern blot analysis. The probes for 1b gene (nt 18057-18222; SEQ ID NO:2484), S gene (nt 21920-22107; SEQ ID NO:2485), M gene (nt 25867-26996; SEQ ID NO:2486), and N gene (nt 28658-28883; SEQ ID NO:2487) are shown.

5. DETAILED DESCRIPTION OF THE INVENTION

30 The present inventors developed a rapid, high-throughput reverse transcription-PCR diagnostic test for SARS associated coronavirus (SARS-CoV). 3' region of the hSARS virus genome including the Nucleocapsid gene (N-gene) represents a sensitive molecular

marker which can be used in addition to 1b gene to increase the sensitivity of the test. An internal control using PK-15 cells may be employed to ensure the integrity of RNA during its extraction process and cDNA synthesis, thus eliminating false negative results.

In mouse hepatitis virus (MHV), a typical member of the genus Coronavirus, both genomic RNA and mRNA transcripts are capped and with common 3' ends and common leader sequences on their 5' ends. With this unique transcription strategy, the copy numbers of different viral genes during proliferation of virus in its host are different (Figure 19). N gene that encodes for the nucleocapsid has the most abundant copy number during virus replication as all transcripts may carry nucleotide sequence from N gene, although they are not all in-frame for translation for this gene product. The present inventors have discovered a diagnostic assay that is based on the 3' region, including the N-gene, of the viral genome provide a more sensitive assay than the rest of the viral genome. Accordingly, in preferred embodiments, nucleic acid molecules that may be used for a diagnostic assay comprise nucleic acid sequence of nucleotide position 18000 to 29742 of SEQ ID NO:15, or portions thereof. The portions may comprise 15, 20, 25, 30 , 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200 of nucleotides having nucleic acid sequence from nucleotide position 18000 to 29742 of SEQ ID NO:15. In other preferred embodiments, nucleic acid molecules that may be used for a diagnostic assay comprising nucleic acid sequence of nucleotide position 28658 to 28883 or 29247-29410 of SEQ ID NO:15.

Nasopharyngeal aspirate (NPA) and stool samples were obtained from SARS suspected patients with major clinical symptoms and significant history of close contact with infected patients. Total RNA was extracted from the subject samples, together with PK-15 cell as an internal control. Samples were analyzed by the reverse-transcription-PCR assay. Northern blot analysis was performed to show different subgenomic transcripts of the virus. Real-time quantitative PCR was employed to compare the sensitivity of two loci used in this diagnostic assay. In specific embodiments, PCR inhibitor was removed with ethanol precipitation after RNA extraction process.

In preferred embodiments, the present invention provides a method for detecting the presence or absence of nucleic acid of the N-gene in a biological sample. The method involves obtaining a biological sample from various sources and contacting the sample with a compound or an agent capable of detecting a nucleic acid (e.g., mRNA, genomic RNA) of the N-gene of the hSARS virus such that the presence of the N-gene is detected in the

sample. In preferred embodiments, the N-gene may be detected using a labeled nucleic acid probe comprising of the nucleotide sequence of SEQ ID NO:2471, complement thereof, or portion thereof. The portion may be 10, 20, 30, 40, 50, 100, 200, 400, 500, 600, 800, 1000, 1200 nucleotides in length. In preferred embodiments, primers comprising nucleotide sequence of SEQ ID NOS:2475 and/or 2476 or SEQ ID NOS:2480 and/or 2481 may be used to amplify a portion of the N-gene for detection.

5 A preferred agent for detecting hSARS mRNA or genomic RNA of the invention is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic RNA encoding a polypeptide of the invention. The nucleic acid probe can be, for example, a nucleic acid molecule comprising or consisting of the nucleotide sequence or SEQ ID NO:1, 11, 13, 15, 10 2471 or 2473, complement thereof, or a portion thereof, such as an oligonucleotide of at least 15, 20, 25, 30, 50, 100, 250, 500, 750, 1,000 or more contiguous nucleotides in length and sufficient to specifically hybridize under stringent conditions to a hSARS mRNA or genomic RNA.

15 In another preferred specific embodiment, the presence of N-gene is detected in the sample by an reverse transcription polymerase chain reaction (RT-PCR) using the primers that are constructed based on a partial nucleotide sequence of the N-gene or a genomic nucleic acid sequence of SEQ ID NO:15, or based on a nucleotide sequence of SEQ ID NO:1, 11, 13, 15, 2471, or 2473. In a non-limiting specific embodiment, preferred primers 20 to be used in a RT-PCR method are: 5'-TACACACCTCAGC-GTTG-3' (SEQ ID NO:3) and/or 5'-CACGAACGTGACG-AAT-3' (SEQ ID NO:4), in the presence of 2.5 mM MgCl₂ and the thermal cycles are, for example, but not limited to, 94 °C for 8 min followed by 40 cycles of 94 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min (*also see Sections 6.7 and 6.8 infra*). In preferred embodiments, the primers comprise nucleic acid sequence of 25 SEQ ID NOS:2475 and 2476. In another preferred embodiment, the primers comprise nucleic acid sequence of SEQ ID NOS:2480 and 2481. In preferred embodiments, the thermal cycles are 94 °C for 10 min followed by 40 cycles of 94 °C for 30 seconds, 56 °C for 30 seconds, 72 °C for 30 seconds, 72°C for 10 minutes. In another preferred embodiment, the thermal cycles are 94 °C for 3 min followed by 40 cycles of 94 °C for 30 30 seconds, 56 °C for 30 seconds, 72 °C for 30 seconds, 72°C for 10 minutes. In more preferred specific embodiment, the present invention provides a real-time quantitative PCR assay to detect the presence of hSARS virus in a biological sample by subjecting the cDNA

obtained by reverse transcription of the extracted total RNA from the sample to PCR reactions using the specific primers, such as those having nucleotide sequences of SEQ ID NOS:3 and 4, and a fluorescence dye, such as SYBR® Green I, which fluoresces when bound non-specifically to double-stranded DNA. The fluorescence signals from these 5 reactions are captured at the end of extension steps as PCR product is generated over a range of the thermal cycles, thereby allowing the quantitative determination of the viral load in the sample based on an amplification plot (*see* Section 6.7, *infra*).

In the preferred embodiment, the present invention provides a method for detecting the presence or absence of nucleic acid of the S-gene in a biological sample. The method 10 involves obtaining a biological sample from various sources and contacting the sample with a compound or an agent capable of detecting a nucleic acid (e.g., mRNA, genomic RNA) of the S-gene of the hSARS virus such that the presence of the S-gene is detected in the sample. A preferred agent for detecting hSARS mRNA or genomic RNA of the invention is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic RNA encoding a 15 polypeptide of the invention. The nucleic acid probe can be, for example, a nucleic acid molecule comprising or consisting of the nucleotide sequence or SEQ ID NO:1, 11, 13, 15, 2471, or 2473, or a portion thereof, such as an oligonucleotide of at least 15, 20, 25, 30, 50, 100, 250, 500, 750, 1,000 or more contiguous nucleotides in length and sufficient to specifically hybridize under stringent conditions to a hSARS mRNA or genomic RNA.

20 In another preferred specific embodiment, the presence of S-gene is detected in the sample by an reverse transcription polymerase chain reaction (RT-PCR) using the primers that are constructed based on a partial nucleotide sequence of the S-gene (SEQ ID NO:2473).

25 In vitro techniques for detection of mRNA include northern hybridizations, *in situ* hybridizations, RT-PCR, and RNase protection. In vitro techniques for detection of genomic RNA include nothern hybridizations, RT-PCT, and RNase protection.

The polynucleotides encoding the N-gene may be amplified before they are detected. The term “amplified” refers to the process of making multiple copies of the nucleic acid 30 from a single polynucleotide molecule. The amplification of polynucleotides can be carried out *in vitro* by biochemical processes known to those of skill in the art. The amplification agent may be any compound or system that will function to accomplish the synthesis of primer extension products, including enzymes. Suitable enzymes for this purpose include, for example, *E. coli* DNA polymerase I, Taq polymerase, Klenow fragment of *E. coli* DNA

polymerase I, T4 DNA polymerase, other available DNA polymerases, polymerase muteins, reverse transcriptase, ligase, and other enzymes, including heat-stable enzymes (*i.e.*, those enzymes that perform primer extension after being subjected to temperatures sufficiently elevated to cause denaturation). Suitable enzymes will facilitate combination of the

- 5 nucleotides in the proper manner to form the primer extension products that are complementary to each mutant nucleotide strand. Generally, the synthesis will be initiated at the 3'-end of each primer and proceed in the 5'-direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be amplification agents, however, that initiate synthesis at the 5'-end and proceed in the other direction,
10 using the same process as described above. In any event, the method of the invention is not to be limited to the embodiments of amplification described herein.

One method of *in vitro* amplification, which can be used according to this invention, is the polymerase chain reaction (PCR) described in U.S. Patent Nos. 4,683,202 and 4,683,195. The term "polymerase chain reaction" refers to a method for amplifying a DNA
15 base sequence using a heat-stable DNA polymerase and two oligonucleotide primers, one complementary to the (+)-strand at one end of the sequence to be amplified and the other complementary to the (-)-strand at the other end. Because the newly synthesized DNA strands can subsequently serve as additional templates for the same primer sequences, successive rounds of primer annealing, strand elongation, and dissociation produce rapid
20 and highly specific amplification of the desired sequence. The polymerase chain reaction is used to detect the presence of polynucleotides encoding cytokines in the sample. Many polymerase chain methods are known to those of skill in the art and may be used in the method of the invention. For example, DNA can be subjected to 30 to 35 cycles of amplification in a thermocycler as follows: 95°C for 30 sec, 52° to 60°C for 1 min, and
25 72°C for 1 min, with a final extension step of 72°C for 5 min. For another example, DNA can be subjected to 35 polymerase chain reaction cycles in a thermocycler at a denaturing temperature of 95°C for 30 sec, followed by varying annealing temperatures ranging from 54-58°C for 1 min, an extension step at 70°C for 1 min and a final extension step at 70°C.

The primers for use in amplifying the N-gene or S-gene of the invention may be
30 prepared using any suitable method, such as conventional phosphotriester and phosphodiester methods or automated embodiments thereof so long as the primers are capable of hybridizing to the polynucleotides of interest. One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Patent No. 4,458,066. The

exact length of primer will depend on many factors, including temperature, buffer, and nucleotide composition. The primer must prime the synthesis of extension products in the presence of the inducing agent for amplification.

Primers used according to the method of the invention are complementary to each strand of nucleotide sequence to be amplified. The term "complementary" means that the primers must hybridize with their respective strands under conditions, which allow the agent for polymerization to function. In other words, the primers that are complementary to the flanking sequences hybridize with the flanking sequences and permit amplification of the nucleotide sequence. Preferably, the 3' terminus of the primer that is extended has perfectly base paired complementarity with the complementary flanking strand. Primers and probes for polynucleotides encoding N-gene or S-gene of the present invention can be developed using known methods combined with the present disclosure.

Those of ordinary skill in the art will know of various amplification methodologies that can also be utilized to increase the copy number of target nucleic acid. The polynucleotides detected in the method of the invention can be further evaluated, detected, cloned, sequenced, and the like, either in solution or after binding to a solid support, by any method usually applied to the detection of a specific nucleic acid sequence such as another polymerase chain reaction, oligomer restriction (Saiki *et al.*, *BioTechnology* 3:1008-1012 (1985)), allele-specific oligonucleotide (ASO) probe analysis (Conner *et al.*, *Proc. Natl. Acad. Sci. USA* 80: 278 (1983), oligonucleotide ligation assays (OLAs) (Landegren *et al.*, *Science* 241:1077 (1988)), RNase Protection Assay and the like. Molecular techniques for DNA analysis have been reviewed (Landegren *et al.*, *Science* 242: 229-237 (1988)). Following DNA amplification, the reaction product may be detected by Southern blot analysis, without using radioactive probes. In such a process, for example, a small sample of DNA containing the polynucleotides obtained from the tissue or subject are amplified, and analyzed via a Southern blotting technique. The use of non-radioactive probes or labels is facilitated by the high level of the amplified signal. In one embodiment of the invention, one nucleoside triphosphate is radioactively labeled, thereby allowing direct visualization of the amplification product by autoradiography. In another embodiment, amplification primers are fluorescently labeled and run through an electrophoresis system. Visualization of amplified products is by laser detection followed by computer assisted graphic display, without a radioactive signal.

The methods of the present invention can involve a real-time quantitative PCR assay, such as a Taqman® assay (Holland et al., *Proc Natl Acad Sci U.S.A.*, 88(16):7276 (1991); also see U.S. Patent Application of Attorney Docket No. V9661.0078 filed March 24, 2004, which is incorporated by reference in its entirety). The assays can be performed on an instrument designed to perform such assays, for example those available from Applied Biosystems (Foster City, CA). Primers and probes for such an assay can be designed according to known procedures in the art.

5 The size of the primers used to amplify a portion of the N-gene or S-gene is at least 10, 15, 20, 25, 30 nucleotide in length. In particular, primers that amplify the N-gene or S- gene is most preferred. Preferably, the GC ratio should be above 30, 35, 40, 45, 50, 55, 60 10 % so as to prevent hair-pin structure on the primer. Furthermore, the amplicon should be sufficiently long enough to be detected by standard molecular biology methodologies. Preferably, the amplicon is at least 40, 60, 100, 200, 300, 400, 500, 600, 800, 1000 base pair in length.

15 In a specific embodiment, the methods further involve obtaining a control sample from a control subject, contacting the control sample with a compound or agent capable of detecting N-gene or S-gene, such that the presence of mRNA or genomic RNA encoding the N-gene or S-gene is detected in the sample, and comparing the presence (or absence) of N- gene or S-gene, or mRNA or genomic RNA encoding the polypeptide in the control sample 20 with the presence of N-gene or S-gene, or mRNA or genomic DNA encoding the polypeptide in the test sample.

The invention also encompasses kits for detecting the presence of N-gene nucleic acid in a test sample. The kit, for example, can comprise a labeled compound or agent capable of detecting a nucleic acid molecule encoding the polypeptide in a test sample and, 25 in certain embodiments, a means for determining the amount of mRNA in the sample (an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide).

For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide of the invention or to a sequence within the N-gene; 30 (2) a pair of primers useful for amplifying a nucleic acid molecule containing the N-gene sequence. The kit can also comprise, e.g., a buffering agent, a preservative, or a protein stabilizing agent. The kit can also comprise components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample

or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with instructions for use.

The present invention relates to the isolated N-gene and S-gene of the hSARS virus.

- 5 In a specific embodiment, the virus comprises a nucleotide sequence of SEQ ID NO:1, 11, 13, 15, 2471, and/or 2473. In a specific embodiment, the present invention provides isolated nucleic acid molecules of the hSARS virus, comprising, or, alternatively, consisting of the nucleotide sequence of SEQ ID NO:1, 11, 13, 15, 2471, and/or, 2473, a complement thereof or a portion thereof. In another specific embodiment, the invention provides
- 10 isolated nucleic acid molecules which hybridize under stringent conditions, as defined herein, to a nucleic acid molecule having the sequence of SEQ ID NO:1, 11, 13, 15, 2471, and/or 2473, or specific genes of known member of *Coronaviridae*, or a complement thereof. In another specific embodiment, the invention provides isolated polypeptides or proteins that are encoded by a nucleic acid molecule comprising a nucleotide sequence that
- 15 is at least about 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:1, or a complement thereof. In another specific embodiment, the invention provides isolated polypeptides or proteins that are encoded by a nucleic acid molecule comprising a nucleotide sequence that is at least about 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200,
- 20 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200, or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:11, or a complement thereof. In yet another specific embodiment, the invention provides isolated polypeptides or proteins that are encoded by a nucleic acid molecule comprising a nucleotide sequence that is at least about 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200,
- 25 300, 350, 400, 450, 500, 550, 600, 650, 700, or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:13, or a complement thereof. In yet another specific embodiment, the invention provides isolated polypeptides or proteins that are encoded by a nucleic acid molecule comprising or, alternatively consisting of a nucleotide sequence that is at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600,
- 30 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200 or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2471, or a complement thereof. In yet another specific embodiment, the invention provides isolated polypeptides or proteins that are encoded by a nucleic acid molecule comprising or, alternatively consisting of a

nucleotide sequence that is at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200, 2,000, 3,000, or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2473, or a complement thereof. In yet another specific embodiment, the invention

- 5 provides isolated polypeptides or proteins that are encoded by a nucleic acid molecule comprising or, alternatively consisting of a nucleotide sequence that is at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000,
10 20,000, 21,000, 22,000, 23,000, 24,000, 25,000, 26,000, 27,000, 28,000, 29,000 or more contiguous nucleotides of the nucleotide sequence of SEQ ID NO:15, or a complement thereof. The polypeptides includes those shown in Figures 11 (SEQ ID NOS:17-239, 241-736 and 738-1107) and 12 (SEQ ID NOS:1109-1589, 1591-1964 and 1966-2470) or having an amino acid sequence of SEQ ID NO:2472 or 2474. The polypeptides or the proteins of
15 the present invention preferably have one or more biological activities of the proteins encoded by the sequence of SEQ ID NO:1, 11, 13, 15, 2471 or 2473, or the polypeptides shown in Figures 11 and 12, or the native viral proteins containing the amino acid sequences encoded by the sequence of SEQ ID NO:1, 11, 13, 15, 2471 or 2473.

The present invention also relates to a method for propagating the hSARS virus in
20 host cells.

The invention further relates to the use of the sequence information of the isolated virus for diagnostic and therapeutic methods. In a specific embodiment, the invention provides the entire nucleotide sequence of hSARS virus, CCTCC-V200303, SEQ ID NO:15, or fragments, or complement thereof. Furthermore, the present invention relates to a
25 nucleic acid molecule that hybridizes any portion of the genome of the hSARS virus, CCTCC-V200303, or SEQ ID NO:15, under the stringent conditions. In a specific embodiment, the invention provides nucleic acid molecules which are suitable for use as primers consisting of or comprising the nucleotide sequence of SEQ ID NO:1, 11, 13, 15, 2471 or 2473, or a complement thereof, or a portion thereof. In specific embodiments, the primers comprise nucleotide sequence of SEQ ID NO: 2475, 2476, 2477, 2478, 2480 or 2481. In another specific embodiment, the invention provides nucleic acid molecules which are suitable for use as hybridization probes for the detection of nucleic acids encoding a polypeptide of the invention, consisting of or comprising the nucleotide sequence of SEQ

ID NO:1, 11, 13, 15, 2471 or 2473, a complement thereof, or a portion thereof. The invention further relates to a kit comprising primers having nucleic acid sequence of SEQ ID NOS:2475 and 2476; and SEQ ID NOS:2480 and 2481, for the detection of N-gene. In another embodiment, the invention relates to a kit comprising primers having nucleic acid sequence of SEQ ID NOS:2477 and/or 2478 for the detection of S-gene. In a preferred embodiment, the kit further comprises reagents for the detection of genes not found in hSARS virus as a negative control. The invention further encompasses chimeric or recombinant viruses or viral proteins encoded by said nucleotide sequences.

The invention further provides antibodies that specifically bind a polypeptide of the invention encoded by the nucleotide sequence of SEQ ID NO:1, 11, 13, 2471 or 2473, or a fragment thereof, or any hSARS epitope. The invention further provides antibodies that specifically bind a polypeptide having amino acid sequence of SEQ ID NO: 2472 or 2474. The invention further provides antibodies that specifically bind the polypeptides of the invention encoded by the nucleotide sequence of SEQ ID NO:15, or the polypeptides shown in Figures 11 and 12, or a fragment thereof, or any hSARS epitope. Such antibodies include, but are not limited to polyclonal, monoclonal, bi-specific, multi-specific, human, humanized, chimeric antibodies, single chain antibodies, Fab fragments, F(ab')₂ fragments, disulfide-linked Fvs, intrabodies and fragments containing either a VI or VH domain or even a complementary determining region (CDR) that specifically binds to a polypeptide of the invention.

In one embodiment, the invention provides methods for detecting the presence, activity or expression of the hSARS virus of the invention in a biological material, such as cells, blood, saliva, urine, sputum, nasopharyngeal aspirates, and so forth. The presence of the hSARS virus in a sample can be determined by contacting the biological material with an agent which can detect directly or indirectly the presence of the hSARS virus. In a specific embodiment, the detection agents are the antibodies of the present invention. In another embodiment, the detection agent is a nucleic acid of the present invention.

In another embodiment, the invention provides vaccine preparations comprising the hSARS virus, including recombinant and chimeric forms of said virus, or subunits of the virus. In a specific embodiment, the vaccine preparations comprise live but attenuated hSARS virus with or without pharmaceutically acceptable excipients, including adjuvants. In another specific embodiment, the vaccine preparations comprise an inactivated or killed hSARS virus with or without pharmaceutically acceptable excipients, including adjuvants.

The vaccine preparations of the present invention may further comprise with adjuvants or

Accordingly, the present invention further provides methods of preparing recombinant or chimeric forms of hSARS. In another specific invention, the vaccine preparations of the present invention comprise one or more nucleic acid molecules comprising or consisting of

5 the sequence of SEQ ID NO. 1, 11, 13, 15, 2471 and/or 2473, or a fragment thereof. In another embodiment, the invention provides vaccine preparations comprising one or more polypeptides of the invention encoded by a nucleotide sequence comprising or consisting of the nucleotide sequence of SEQ ID NO:1, 11, 13, 2471 and/or 2473, or the polypeptides shown in Figures 11 and 12, or a fragment thereof. In another embodiment, the invention

10 provides vaccine preparations comprising one or more polypeptides of the invention encoded by a nucleotide sequence comprising or consisting of the nucleotide sequence of SEQ ID NO:15, or a fragment thereof. Furthermore, the present invention provides methods for treating, ameliorating, managing, or preventing SARS by administering the vaccine preparations or antibodies of the present invention alone or in combination with

15 antivirals [e.g., amantadine, rimantadine, gancyclovir, acyclovir, ribavirin, penciclovir, oseltamivir, foscarnet zidovudine (AZT), didanosine (ddI), lamivudine (3TC), zalcitabine (ddC), stavudine (d4T), nevirapine, delavirdine, indinavir, ritonavir, vidarabine, nelfinavir, saquinavir, relenza, tamiflu, pleconaril, interferons, etc.], steroids and corticosteroids such as prednisone, cortisone, fluticasone and glucocorticoid, antibiotics, analgesics,

20 bronchodilators, or other treatments for respiratory and/or viral infections.

Furthermore, the present invention provides pharmaceutical compositions comprising anti-viral agents of the present invention and a pharmaceutically acceptable carrier. The present invention also provides kits comprising pharmaceutical compositions of the present invention.

25 In another aspect, the present invention provides methods for screening anti-viral agents that inhibit the infectivity or replication of hSARS virus or variants thereof.

5.1 Recombinant and Chimeric hSARS Viruses

The present invention encompasses recombinant or chimeric viruses encoded by

30 viral vectors derived from the genome of hSARS virus or natural variants thereof. In a specific embodiment, a recombinant virus is one derived from the hSARS virus of deposit accession no. CCTCC-V200303. In a specific embodiment, the virus has a nucleotide

sequence of SEQ ID NO:15. In another specific embodiment, a recombinant virus is one derived from a natural variant of hSARS virus. A natural variant of hSARS has a sequence that is different from the genomic sequence (SEQ ID NO:15) of the hSARS virus, CCTCC-V200303, due to one or more naturally occurred mutations, including, but not limited to,

- 5 point mutations, rearrangements, insertions, deletions etc., to the genomic sequence that may or may not result in a phenotypic change. In accordance with the present invention, a viral vector which is derived from the genome of the hSARS virus, CCTCC-V200303, is one that contains a nucleic acid sequence that encodes at least a part of one ORF of the hSARS virus. In a specific embodiment, the ORF comprises or consists of a nucleotide
10 sequence of SEQ ID NO:1, 11, 13, 2471, 2473, or a fragment thereof. In a specific embodiment, there are more than one ORF within the nucleotide sequence of SEQ ID NO:15, as shown in Figures 11 (SEQ ID NOS:16, 240 and 737) and 12 (SEQ ID NOS:1108, 1590 and 1965), or a fragment thereof. In another embodiment, the polypeptide encoded by the ORF comprises or consists of an amino acid sequence of SEQ ID NO:2, 12, 14, 2472,
15 2474, or a fragment thereof, or shown in Figures 11 (SEQ ID NOS:17-239, 241-736 and 738-1107) and 12 (SEQ ID NOS:1109-1589, 1591-1964 and 1966-2470), or a fragment thereof. In accordance with the present invention these viral vectors may or may not include nucleic acids that are non-native to the viral genome.

In another specific embodiment, a chimeric virus of the invention is a recombinant
20 hSARS virus which further comprises a heterologous nucleotide sequence. In accordance with the invention, a chimeric virus may be encoded by a nucleotide sequence in which heterologous nucleotide sequences have been added to the genome or in which endogenous or native nucleotide sequences have been replaced with heterologous nucleotide sequences.

According to the present invention, the chimeric viruses are encoded by the viral
25 vectors of the invention which further comprise a heterologous nucleotide sequence. In accordance with the present invention a chimeric virus is encoded by a viral vector that may or may not include nucleic acids that are non-native to the viral genome. In accordance with the invention a chimeric virus is encoded by a viral vector to which heterologous nucleotide sequences have been added, inserted or substituted for native or non-native
30 sequences. In accordance with the present invention, the chimeric virus may be encoded by nucleotide sequences derived from different strains or variants of hSARS virus. In particular, the chimeric virus is encoded by nucleotide sequences that encode antigenic polypeptides derived from different strains or variants of hSARS virus.

A chimeric virus may be of particular use for the generation of recombinant vaccines protecting against two or more viruses (Tao et al., J. Virol. 72, 2955-2961; Durbin et al., 2000, J. Virol. 74, 6821-6831; Skiadopoulos et al., 1998, J. Virol. 72, 1762-1768 (1998); Teng et al., 2000, J. Virol. 74, 9317-9321). For example, it can be envisaged that a virus

5 vector derived from the hSARS virus expressing one or more proteins of variants of hSARS virus, or vice versa, will protect a subject vaccinated with such vector against infections by both the native hSARS and the variant. Attenuated and replication-defective viruses may be of use for vaccination purposes with live vaccines as has been suggested for other viruses.

(See, PCT WO 02/057302, at pp. 6 and 23, incorporated by reference herein).

10 In accordance with the present invention the heterologous sequence to be incorporated into the viral vectors encoding the recombinant or chimeric viruses of the invention include sequences obtained or derived from different strains or variants of hSARS.

In certain embodiments, the chimeric or recombinant viruses of the invention are encoded by viral vectors derived from viral genomes wherein one or more sequences, 15 intergenic regions, termini sequences, or portions or entire ORF have been substituted with a heterologous or non-native sequence. In certain embodiments of the invention, the chimeric viruses of the invention are encoded by viral vectors derived from viral genomes wherein one or more heterologous sequences have been inserted or added to the vector.

The selection of the viral vector may depend on the species of the subject that is to 20 be treated or protected from a viral infection. If the subject is human, then an attenuated hSARS virus can be used to provide the antigenic sequences.

In accordance with the present invention, the viral vectors can be engineered to provide antigenic sequences which confer protection against infection by the hSARS and natural variants thereof. The viral vectors may be engineered to provide one, two, three or 25 more antigenic sequences. In accordance with the present invention the antigenic sequences may be derived from the same virus, from different strains or variants of the same type of virus, or from different viruses.

The expression products and/or recombinant or chimeric virions obtained in accordance with the invention may advantageously be utilized in vaccine formulations. The 30 expression products and chimeric virions of the present invention may be engineered to create vaccines against a broad range of pathogens, including viral and bacterial antigens, tumor antigens, allergen antigens, and auto antigens involved in autoimmune disorders. In particular, the chimeric virions of the present invention may be engineered to create

vaccines for the protection of a subject from infections with hSARS virus and variants thereof.

In certain embodiments, the expression products and recombinant or chimeric virions of the present invention may be engineered to create vaccines against a broad range

5 of pathogens, including viral antigens, tumor antigens and autoantigens involved in autoimmune disorders. One way to achieve this goal involves modifying existing hSARS genes to contain foreign sequences in their respective external domains. Where the heterologous sequences are epitopes or antigens of pathogens, these chimeric viruses may be used to induce a protective immune response against the disease agent from which these
10 determinants are derived.

Thus, the present invention relates to the use of viral vectors and recombinant or chimeric viruses to formulate vaccines against a broad range of viruses and/or antigens.

The present invention also encompasses recombinant viruses comprising a viral vector derived from the hSARS or variants thereof which contains sequences which result in a
15 virus having a phenotype more suitable for use in vaccine formulations, e.g., attenuated phenotype or enhanced antigenicity. The mutations and modifications can be in coding regions, in intergenic regions and in the leader and trailer sequences of the virus.

The invention provides a host cell comprising a nucleic acid or a vector according to the invention. Plasmid or viral vectors containing the polymerase components of hSARS
20 virus are generated in prokaryotic cells for the expression of the components in relevant cell types (bacteria, insect cells, eukaryotic cells). Plasmid or viral vectors containing full-length or partial copies of the hSARS genome will be generated in prokaryotic cells for the expression of viral nucleic acids in-vitro or in-vivo. The latter vectors may contain other viral sequences for the generation of chimeric viruses or chimeric virus proteins, may
25 lack parts of the viral genome for the generation of replication defective virus, and may contain mutations, deletions or insertions for the generation of attenuated viruses. In addition, the present invention provides a host cell infected with hSARS virus, for example, of deposit no. CCTCC-V200303.

Infectious copies of hSARS (being wild type, attenuated, replication-defective or
30 chimeric) can be produced upon co-expression of the polymerase components according to the state-of-the-art technologies described above.

In addition, eukaryotic cells, transiently or stably expressing one or more full-length or partial hSARS proteins can be used. Such cells can be made by transfection (proteins or

nucleic acid vectors), infection (viral vectors) or transduction (viral vectors) and may be useful for complementation of mentioned wild type, attenuated, replication-defective or chimeric viruses.

- The viral vectors and chimeric viruses of the present invention may be used to
5 modulate a subject's immune system by stimulating a humoral immune response, a cellular immune response or by stimulating tolerance to an antigen. As used herein, a subject means: humans, primates, horses, cows, sheep, pigs, goats, dogs, cats, avian species and rodents.

5.2 Formulation of Vaccines and Antivirals

- In a preferred embodiment, the invention provides a proteinaceous molecule or hSARS virus specific viral protein or functional fragment thereof encoded by a nucleic acid according to the invention. Useful proteinaceous molecules are for example derived from any of the genes or genomic fragments derivable from the virus according to the invention, including envelop protein (E protein), integral membrane protein (M protein), spike protein 10 (S protein), nucleocapsid protein (N protein), hemagglutinin esterase (HE protein), and RNA-dependent RNA polymerase. Such molecules, or antigenic fragments thereof, as provided herein, are for example useful in diagnostic methods or kits and in pharmaceutical compositions such as subunit vaccines. Particularly useful are polypeptides encoded by the nucleotide sequence of SEQ ID NO:1, 11, 13, 15, 2471, 2473, or as shown in Fig. 11 (SEQ 15 ID NOS:17-239, 241-736 and 738-1107) and Fig. 12 (SEQ ID NOS:1109-1589, 1591-1964 and 1966-2470), or having the amino acid sequence of SEQ ID NO:2472 or 2474, or antigenic fragments thereof for inclusion as antigen or subunit immunogen, but inactivated whole virus can also be used. Particularly useful are also those proteinaceous substances 20 that are encoded by recombinant nucleic acid fragments of the hSARS genome, of course preferred are those that are within the preferred bounds and metes of ORFs, in particular, for eliciting hSARS specific antibody or T cell responses, whether *in vivo* (e.g. for protective or therapeutic purposes or for providing diagnostic antibodies) or *in vitro* (e.g. by phage 25 display technology or another technique useful for generating synthetic antibodies).

- The invention provides vaccine formulations for the prevention and treatment of infections with hSARS virus. In certain embodiments, the vaccine of the invention 30 comprises recombinant and chimeric viruses of the hSARS virus. In certain embodiments, the virus is attenuated.

In another embodiment of this aspect of the invention, inactivated vaccine formulations may be prepared using conventional techniques to "kill" the chimeric viruses. Inactivated vaccines are "dead" in the sense that their infectivity has been destroyed. Ideally, the infectivity of the virus is destroyed without affecting its immunogenicity. In 5 order to prepare inactivated vaccines, the chimeric virus may be grown in cell culture or in the allantois of the chick embryo, purified by zonal ultracentrifugation, inactivated by formaldehyde or β -propiolactone, and pooled. The resulting vaccine is usually inoculated intramuscularly.

Inactivated viruses may be formulated with a suitable adjuvant in order to enhance 10 the immunological response. Such adjuvants may include but are not limited to mineral gels, e.g., aluminum hydroxide; surface active substances such as lyssolecithin, pluronic polyols, polyanions; peptides; oil emulsions; and potentially useful human adjuvants such as BCG and *Corynebacterium parvum*.

In another aspect, the present invention also provides DNA vaccine formulations 15 comprising a nucleic acid or fragment of the hSARS virus, e.g., the virus having accession no. CCTCC-V200303, or nucleic acid molecules having the sequence of SEQ ID NO:1, 11, 13, 15, 2471, 2473, or a fragment thereof. In another specific embodiment, the DNA 20 vaccine formulations of the present invention comprises a nucleic acid or fragment thereof encoding the antibodies which immunospecifically binds hSARS viruses. In DNA vaccine formulations, a vaccine DNA comprises a viral vector, such as that derived from the hSARS virus, bacterial plasmid, or other expression vector, bearing an insert comprising a nucleic acid molecule of the present invention operably linked to one or more control elements, thereby allowing expression of the vaccinating proteins encoded by said nucleic acid 25 molecule in a vaccinated subject. Such vectors can be prepared by recombinant DNA technology as recombinant or chimeric viral vectors carrying a nucleic acid molecule of the present invention (*see also* Section 5.1, *supra*).

Various heterologous vectors are described for DNA vaccinations against viral infections. For example, the vectors described in the following references may be used to express hSARS sequences instead of the sequences of the viruses or other pathogens 30 described; in particular, vectors described for hepatitis B virus (Michel, M.L. *et al.*, 1995, DAN-mediated immunization to the hepatitis B surface antigen in mice: Aspects of the humoral response mimic hepatitis B viral infection in humans, *Proc. Natl. Acad. Sci. USA* 92:5307-5311; Davis, H.L. *et al.*, 1993, DNA-based immunization induces continuous

- seretion of hepatitis B surface antigen and high levels of circulating antibody, *Human Mol. Genetics* 2:1847-1851), HIV virus (Wang, B. *et al.*, 1993, Gene inoculation generates immune responses against human immunodeficiency virus type 1, *Proc. Natl. Acad. Sci. USA* 90:4156-4160; Lu, S. *et al.*, 1996, Simian immunodeficiency virus DNA vaccine trial in macques, *J. Virol.* 70:3978-3991; Letvin, N.L. *et al.*, 1997, Potent, protective anti-HIV immune responses generated by bimodal HIV envelope DNA plus protein vaccination, *Proc. Natl. Acad. Sci. U.S.A.* 94(17):9378-83), and influenza viruses (Robinson, Hl *et al.*, 1993, Protection against a lethal influenza virus challenge by immunization with a haemagglutinin-expressing plasmid DNA, *Vaccine* 11:957-960; Ulmer, J.B. *et al.*,
- 5 Heterologous protection against influenza by injection of DNA encoding a viral protein, *Science* 259:1745-1749), as well as bacterial infections, such as tuberculosis (Tascon, R.E. *et al.*, 1996, Vaccination against tuberculosis by DNA injection, *Nature Med.* 2:888-892; Huygen, K. *et al.*, 1996, Immunogenicity and protective efficacy of a tuberculosis DNA vaccine, *Nature Med.*, 2:893-898), and parasitic infection, such as malaria (Sedegah, M., 10 1994, Protection against malaria by immunization with plasmid DNA encoding circumsporozoite protein, *Proc. Natl. Acad. Sci. USA* 91:9866-9870; Doolan, D.L. *et al.*, 1996, Circumventing genetic restriction of protection against malaria with multigene DNA immunization: CD8+ T cell-interferon δ , and nitric oxide-dependent immunity, *J. Exper. Med.*, 1183:1739-1746).
- 15 20 Many methods may be used to introduce the vaccine formulations described above. These include, but are not limited to, oral, intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, and intranasal routes. Alternatively, it may be preferable to introduce the chimeric virus vaccine formulation via the natural route of infection of the pathogen for which the vaccine is designed. The DNA vaccines of the present invention 25 may be administered in saline solutions by injections into muscle or skin using a syringe and needle (Wolff J.A. *et al.*, 1990, Direct gene transfer into mouse muscle in vivo, *Science* 247:1465-1468; Raz, E., 1994, Intradermal gene immunization: The possible role of DNA uptake in the induction of cellular immunity to viruses, *Proc. Natl. Acad. Sci. USA* 91:9519-9523). Another way to administer DNA vaccines is called "gene gun" method, whereby 30 microscopic gold beads coated with the DNA molecules of interest is fired into the cells (Tang, D. *et al.*, 1992, Genetic immunization is a simple method for eliciting an immune response, *Nature* 356:152-154). For general reviews of the methods for DNA vaccines, see Robinson, H.L., 1999, DNA vaccines: basic mechanism and immune responses (Review),

Int. J. Mol. Med. 4(5):549-555; Barber, B., 1997, Introduction: Emerging vaccine strategies, *Seminars in Immunology* 9(5):269-270; and Robinson, H.L. *et al.*, 1997, DNA vaccines, *Seminars in Immunology* 9(5):271-283.

5 **5.3 Attenuation of hSARS Virus or Variants Thereof**

The hSARS virus or variants thereof of the invention can be genetically engineered to exhibit an attenuated phenotype. In particular, the viruses of the invention exhibit an attenuated phenotype in a subject to which the virus is administered as a vaccine.

Attenuation can be achieved by any method known to a skilled artisan. Without being

10 bound by theory, the attenuated phenotype of the viruses of the invention can be caused, *e.g.*, by using a virus that naturally does not replicate well in an intended host species, for example, by reduced replication of the viral genome, by reduced ability of the virus to infect a host cell, or by reduced ability of the viral proteins to assemble to an infectious viral particle relative to the wild type strain of the virus.

15 The attenuated phenotypes of hSARS virus or variants thereof can be tested by any method known to the artisan. A candidate virus can, for example, be tested for its ability to infect a host or for the rate of replication in a cell culture system. In certain embodiments, growth curves at different temperatures are used to test the attenuated phenotype of the virus. For example, an attenuated virus is able to grow at 35°C, but not at 39°C or 40°C. In
20 certain embodiments, different cell lines can be used to evaluate the attenuated phenotype of the virus. For example, an attenuated virus may only be able to grow in monkey cell lines but not the human cell lines, or the achievable virus titers in different cell lines are different for the attenuated virus. In certain embodiments, viral replication in the respiratory tract of a small animal model, including but not limited to, hamsters, cotton rats, mice and guinea
25 pigs, is used to evaluate the attenuated phenotypes of the virus. In other embodiments, the immune response induced by the virus, including but not limited to, the antibody titers (*e.g.*, assayed by plaque reduction neutralization assay or ELISA) is used to evaluate the attenuated phenotypes of the virus. In a specific embodiment, the plaque reduction neutralization assay or ELISA is carried out at a low dose. In certain embodiments, the
30 ability of the hSARS virus to elicit pathological symptoms in an animal model can be tested. A reduced ability of the virus to elicit pathological symptoms in an animal model system is

indicative of its attenuated phenotype. In a specific embodiment, the candidate viruses are tested in a monkey model for nasal infection, indicated by mucous production.

The viruses of the invention can be attenuated such that one or more of the functional characteristics of the virus are impaired. In certain embodiments, attenuation is measured in comparison to the wild type strain of the virus from which the attenuated virus is derived. In other embodiments, attenuation is determined by comparing the growth of an attenuated virus in different host systems. Thus, for a non-limiting example, hSARS virus or a variant thereof is said to be attenuated when grown in a human host if the growth of the hSARS or variant thereof in the human host is reduced compared to the non-attenuated hSARS or variant thereof.

In certain embodiments, the attenuated virus of the invention is capable of infecting a host, is capable of replicating in a host such that infectious viral particles are produced. In comparison to the wild type strain, however, the attenuated strain grows to lower titers or grows more slowly. Any technique known to the skilled artisan can be used to determine the growth curve of the attenuated virus and compare it to the growth curve of the wild type virus.

In certain embodiments, the attenuated virus of the invention (*e.g.*, a recombinant or chimeric hSARS) cannot replicate in human cells as well as the wild type virus (*e.g.*, wild type hSARS) does. However, the attenuated virus can replicate well in a cell line that lack interferon functions, such as Vero cells.

In other embodiments, the attenuated virus of the invention is capable of infecting a host, of replicating in the host, and of causing proteins of the virus of the invention to be inserted into the cytoplasmic membrane, but the attenuated virus does not cause the host to produce new infectious viral particles. In certain embodiments, the attenuated virus infects the host, replicates in the host, and causes viral proteins to be inserted in the cytoplasmic membrane of the host with the same efficiency as the wild type hSARS. In other embodiments, the ability of the attenuated virus to cause viral proteins to be inserted into the cytoplasmic membrane into the host cell is reduced compared to the wild type virus. In certain embodiments, the ability of the attenuated hSARS virus to replicate in the host is reduced compared to the wild type virus. Any technique known to the skilled artisan can be used to determine whether a virus is capable of infecting a mammalian cell, of replicating within the host, and of causing viral proteins to be inserted into the cytoplasmic membrane of the host.

In certain embodiments, the attenuated virus of the invention is capable of infecting a host. In contrast to the wild type hSARS, however, the attenuated hSARS cannot be replicated in the host. In a specific embodiment, the attenuated hSARS virus can infect a host and can cause the host to insert viral proteins in its cytoplasmic membranes, but the attenuated virus is incapable of being replicated in the host. Any method known to the skilled artisan can be used to test whether the attenuated hSARS has infected the host and has caused the host to insert viral proteins in its cytoplasmic membranes.

5 In certain embodiments, the ability of the attenuated virus to infect a host is reduced compared to the ability of the wild type virus to infect the same host. Any technique known to the skilled artisan can be used to determine whether a virus is capable of infecting a host.

10 In certain embodiments, mutations (*e.g.*, missense mutations) are introduced into the genome of the virus, for example, into the sequence of SEQ ID NO:1, 11, 13, 15, 2471 or 2473, or to generate a virus with an attenuated phenotype. Mutations (*e.g.*, missense mutations) can be introduced into the structural genes and/or regulatory genes of the hSARS.

15 Mutations can be additions, substitutions, deletions, or combinations thereof. Such variant of hSARS can be screened for a predicted functionality, such as infectivity, replication ability, protein synthesis ability, assembling ability, as well as cytopathic effect in cell cultures. In a specific embodiment, the missense mutation is a cold-sensitive mutation. In another embodiment, the missense mutation is a heat-sensitive mutation. In another embodiment, the missense mutation prevents a normal processing or cleavage of the viral proteins.

20 In other embodiments, deletions are introduced into the genome of the hSARS virus, which result in the attenuation of the virus.

25 In certain embodiments, attenuation of the virus is achieved by replacing a gene of the wild type virus with a gene of a virus of a different species, of a different subgroup, or of a different variant. In another aspect, attenuation of the virus is achieved by replacing one or more specific domains of a protein of the wild type virus with domains derived from the corresponding protein of a virus of a different species. In certain other embodiments, attenuation of the virus is achieved by deleting one or more specific domains of a protein of the wild type virus.

30 When a live attenuated vaccine is used, its safety must also be considered. The vaccine must not cause disease. Any techniques known in the art that can make a vaccine safe may be used in the present invention. In addition to attenuation techniques, other

techniques may be used. One non-limiting example is to use a soluble heterologous gene that cannot be incorporated into the virion membrane. For example, a single copy of the soluble version of a viral transmembrane protein lacking the transmembrane and cytosolic domains thereof, can be used.

- 5 Various assays can be used to test the safety of a vaccine. For example, sucrose gradients and neutralization assays can be used to test the safety. A sucrose gradient assay can be used to determine whether a heterologous protein is inserted in a virion. If the heterologous protein is inserted in the virion, the virion should be tested for its ability to cause symptoms in an appropriate animal model since the virus may have acquired new, 10 possibly pathological, properties.

5.4 Adjuvants and Carrier Molecules

hSARS-associated antigens are administered with one or more adjuvants. In one embodiment, the hSARS-associated antigen is administered together with a mineral salt 15 adjuvants or mineral salt gel adjuvant. Such mineral salt and mineral salt gel adjuvants include, but are not limited to, aluminum hydroxide (ALHYDROGEL, REHYDRAGEL), aluminum phosphate gel, aluminum hydroxyphosphate (ADJU-PHOS), and calcium phosphate.

- In another embodiment, hSARS-associated antigen is administered with an 20 immunostimulatory adjuvant. Such class of adjuvants, include, but are not limited to, cytokines (e.g., interleukin-2, interleukin-7, interleukin-12, granulocyte-macrophage colony stimulating factor (GM-CSF), interferon- γ interleukin-1 β (IL-1 β), and IL-1 β peptide or Sclavo Peptide), cytokine-containing liposomes, triterpenoid glycosides or saponins (e.g., QuilA and QS-21, also sold under the trademark STIMULON, ISCOPEP), Muramyl 25 Dipeptide (MDP) derivatives, such as N-acetyl-muramyl-L-threonyl-D-isoglutamine (Threonyl-MDP, sold under the trademark TERMURTIDE), GMDP, N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine, N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxy phosphoryloxy)-ethylamine, muramyl tripeptide phosphatidylethanolamine (MTP-PE), unmethylated CpG dinucleotides and 30 oligonucleotides, such as bacterial DNA and fragments thereof, LPS, monophosphoryl Lipid A (3D-MLA sold under the trademark MPL), and polyphosphazenes.

- In another embodiment, the adjuvant used is a particular adjuvant, including, but not limited to, emulsions, e.g., Freund's Complete Adjuvant, Freund's Incomplete Adjuvant, squalene or squalane oil-in-water adjuvant formulations, such as SAF and MF59, e.g., prepared with block-copolymers, such as L-121 (polyoxypolypropylene/polyoxyethylene) sold under the trademark PLURONIC L-121, Liposomes, Virosomes, coyleates, and immune stimulating complex, which is sold under the trademark ISCOM.
- 5

In another embodiment, a microparticulate adjuvant is used. Microparticulare adjuvants include, but are not limited to biodegradable and biocompatible polyesters, homo- and copolymers of lactic acid (PLA) and glycolic acid (PGA), poly(lactide-co-glycolides) (PLGA) microparticles, polymers that self-associate into particulates (poloxamer particles), soluble polymers (polyphosphazenes), and virus-like particles (VLPs) such as recombinant protein particulates, e.g., hepatitis B surface antigen (HbsAg).

10

Yet another class of adjuvants that may be used include mucosal adjuvants, including but not limited to heat-labile enterotoxin from *Escherichia coli* (LT), cholera holotoxin (CT) and cholera Toxin B Subunit (CTB) from *Vibrio cholerae*, mutant toxins (e.g., LTK63 and LTR72), microparticles, and polymerized liposomes.

15

In other embodiments, any of the above classes of adjuvants may be used in combination with each other or with other adjuvants. For example, non-limiting examples of combination adjuvant preparations that can be used to administer the hSARS-associated antigens of the invention include liposomes containing immunostimulatory protein, cytokines, or T-cell and/or B-cell peptides, or microbes with or without entrapped IL-2 or microparticles containing enterotoxin. Other adjuvants known in the art are also included within the scope of the invention (see *Vaccine Design: The Subunit and Adjuvant Approach*, Chap. 7, Michael F. Powell and Mark J. Newman (eds.), Plenum Press, New York, 1995, 20 which is incorporated herein in its entirety).

25

The effectiveness of an adjuvant may be determined by measuring the induction of antibodies directed against an immunogenic polypeptide containing a hSARS polypeptide epitope, the antibodies resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

30 The polypeptides may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid additional salts (formed with free amino groups of the peptide) and which are formed with inorganic acids, such as, for example, hydrochloric or phosphoric acids, or organic acids such as acetic, oxalic, tartaric, maleic,

and the like. Salts formed with free carboxyl groups may also be derived from inorganic bases, such as, for example, sodium potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine and the like.

5 The vaccines of the invention may be multivalent or univalent. Multivalent vaccines are made from recombinant viruses that direct the expression of more than one antigen.

Many methods may be used to introduce the vaccine formulations of the invention; these include but are not limited to oral, intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal routes, and via scarification (scratching through the 10 top layers of skin, *e.g.*, using a bifurcated needle).

The patient to which the vaccine is administered is preferably a mammal, most preferably a human, but can also be a non-human animal including but not limited to cows, horses, sheep, pigs, fowl (*e.g.*, chickens), goats, cats, dogs, hamsters, mice and rats.

15 5.5 Preparation of Antibodies

Antibodies which specifically recognize a polypeptide of the invention, such as, but not limited to, polypeptides comprising the sequence of SEQ ID NO:2, 12, 14, 2472, 2474, and polypeptides as shown in Figures 11 (SEQ ID NOS:17-239, 241-736 and 738-1107) and 12 (SEQ ID NOS:1109-1589, 1591-1964 and 1966-2470), or hSARS epitope or

20 antigen-binding fragments thereof can be used for detecting, screening, and isolating the polypeptide of the invention or fragments thereof, or similar sequences that might encode similar enzymes from the other organisms. For example, in one specific embodiment, an antibody which immunospecifically binds hSARS epitope, or a fragment thereof, can be used for various *in vitro* detection assays, including enzyme-linked immunosorbent assays

25 (ELISA), radioimmunoassays, Western blot, etc., for the detection of a polypeptide of the invention or, preferably, hSARS, in samples, for example, a biological material, including cells, cell culture media (*e.g.*, bacterial cell culture media, mammalian cell culture media, insect cell culture media, yeast cell culture media, etc.), blood, plasma, serum, tissues, sputum, naseopharyngeal aspirates, etc.

30 Antibodies specific for a polypeptide of the invention or any epitope of hSARS may be generated by any suitable method known in the art. Polyclonal antibodies to an antigen-of-interest, for example, the hSARS virus from deposit no. CCTCC-V200303, or comprises

a nucleotide sequence of SEQ ID NO:15, can be produced by various procedures well known in the art. For example, an antigen can be administered to various host animals including, but not limited to, rabbits, mice, rats, etc., to induce the production of antisera containing polyclonal antibodies specific for the antigen. Various adjuvants may be used to

- 5 increase the immunological response, depending on the host species, and include but are not limited to, Freund's (complete and incomplete) adjuvant, mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanins, dinitrophenol, and potentially useful adjuvants for humans such as BCG (Bacille Calmette-Guerin) and Corynebacterium parvum.
- 10 Such adjuvants are also well known in the art.

Monoclonal antibodies can be prepared using a wide variety of techniques known in the art including the use of hybridoma, recombinant, and phage display technologies, or a combination thereof. For example, monoclonal antibodies can be produced using hybridoma techniques including those known in the art and taught, for example, in Harlow
15 et al., *Antibodies: A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988); Hammerling, et al., in: *Monoclonal Antibodies and T-Cell Hybridomas*, pp. 563-681 (Elsevier, N.Y., 1981) (both of which are incorporated by reference in their entireties). The term "monoclonal antibody" as used herein is not limited to antibodies produced through hybridoma technology. The term "monoclonal antibody" refers to an antibody that is
20 derived from a single clone, including any eukaryotic, prokaryotic, or phage clone, and not the method by which it is produced.

Methods for producing and screening for specific antibodies using hybridoma technology are routine and well known in the art. In a non-limiting example, mice can be immunized with an antigen of interest or a cell expressing such an antigen. Once an
25 immune response is detected, e.g., antibodies specific for the antigen are detected in the mouse serum, the mouse spleen is harvested and splenocytes isolated. The splenocytes are then fused by well known techniques to any suitable myeloma cells. Hybridomas are selected and cloned by limiting dilution. The hybridoma clones are then assayed by methods known in the art for cells that secrete antibodies capable of binding the antigen.
30 Ascites fluid, which generally contains high levels of antibodies, can be generated by inoculating mice intraperitoneally with positive hybridoma clones.

Antibody fragments which recognize specific epitopes may be generated by known techniques. For example, Fab and F(ab')₂ fragments may be produced by proteolytic

cleavage of immunoglobulin molecules, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')₂ fragments). F(ab')₂ fragments contain the complete light chain, and the variable region, the CH1 region and the hinge region of the heavy chain.

- 5 The antibodies of the invention or fragments thereof can be also produced by any method known in the art for the synthesis of antibodies, in particular, by chemical synthesis or preferably, by recombinant expression techniques.

The nucleotide sequence encoding an antibody may be obtained from any information available to those skilled in the art (i.e., from Genbank, the literature, or by 10 routine cloning and sequence analysis). If a clone containing a nucleic acid encoding a particular antibody or an epitope-binding fragment thereof is not available, but the sequence of the antibody molecule or epitope-binding fragment thereof is known, a nucleic acid encoding the immunoglobulin may be chemically synthesized or obtained from a suitable source (e.g., an antibody cDNA library, or a cDNA library generated from, or nucleic acid, 15 preferably poly A+ RNA, isolated from any tissue or cells expressing the antibody, such as hybridoma cells selected to express an antibody) by PCR amplification using synthetic primers hybridizable to the 3' and 5' ends of the sequence or by cloning using an oligonucleotide probe specific for the particular gene sequence to identify, e.g., a cDNA clone from a cDNA library that encodes the antibody. Amplified nucleic acids generated by 20 PCR may then be cloned into replicable cloning vectors using any method well known in the art.

Once the nucleotide sequence of the antibody is determined, the nucleotide sequence of the antibody may be manipulated using methods well known in the art for the manipulation of nucleotide sequences, e.g., recombinant DNA techniques, site directed 25 mutagenesis, PCR, etc. (see, for example, the techniques described in Sambrook et al., *supra*; and Ausubel et al., eds., 1998, *Current Protocols in Molecular Biology*, John Wiley & Sons, NY, which are both incorporated by reference herein in their entireties), to generate antibodies having a different amino acid sequence by, for example, introducing amino acid substitutions, deletions, and/or insertions into the epitope-binding domain regions of the 30 antibodies or any portion of antibodies which may enhance or reduce biological activities of the antibodies.

Recombinant expression of an antibody requires construction of an expression vector containing a nucleotide sequence that encodes the antibody. Once a nucleotide

sequence encoding an antibody molecule or a heavy or light chain of an antibody, or portion thereof has been obtained, the vector for the production of the antibody molecule may be produced by recombinant DNA technology using techniques well known in the art as discussed in the previous sections. Methods which are well known to those skilled in the art

- 5 can be used to construct expression vectors containing antibody coding sequences and appropriate transcriptional and translational control signals. These methods include, for example, *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. The nucleotide sequence encoding the heavy-chain variable region, light-chain variable region, both the heavy-chain and light-chain variable regions, an epitope-
10 binding fragment of the heavy- and/or light-chain variable region, or one or more complementarity determining regions (CDRs) of an antibody may be cloned into such a vector for expression. Thus-prepared expression vector can be then introduced into appropriate host cells for the expression of the antibody. Accordingly, the invention includes host cells containing a polynucleotide encoding an antibody specific for the
15 polypeptides of the invention or fragments thereof.

The host cell may be co-transfected with two expression vectors of the invention, the first vector encoding a heavy chain derived polypeptide and the second vector encoding a light chain derived polypeptide. The two vectors may contain identical selectable markers which enable equal expression of heavy and light chain polypeptides or different selectable
20 markers to ensure maintenance of both plasmids. Alternatively, a single vector may be used which encodes, and is capable of expressing, both heavy and light chain polypeptides. In such situations, the light chain should be placed before the heavy chain to avoid an excess of toxic free heavy chain (Proudfoot, *Nature*, 322:52, 1986; and Kohler, *Proc. Natl. Acad. Sci. USA*, 77:2 197, 1980). The coding sequences for the heavy and light chains may

- 25 comprise cDNA or genomic DNA.

In another embodiment, antibodies can also be generated using various phage display methods known in the art. In phage display methods, functional antibody domains are displayed on the surface of phage particles which carry the polynucleotide sequences encoding them. In a particular embodiment, such phage can be utilized to display antigen
30 binding domains, such as Fab and Fv or disulfide-bond stabilized Fv, expressed from a repertoire or combinatorial antibody library (e.g., human or murine). Phage expressing an antigen binding domain that binds the antigen of interest can be selected or identified with antigen, e.g., using labeled antigen or antigen bound or captured to a solid surface or bead.

Phage used in these methods are typically filamentous phage, including fd and M13. The antigen binding domains are expressed as a recombinantly fused protein to either the phage gene III or gene VIII protein. Examples of phage display methods that can be used to make the immunoglobulins, or fragments thereof, of the present invention include those disclosed

- 5 in Brinkman et al., *J. Immunol. Methods*, 182:41-50, 1995; Ames et al., *J. Immunol. Methods*, 184:177-186, 1995; Kettleborough et al., *Eur. J. Immunol.*, 24:952-958, 1994; Persic et al., *Gene*, 187:9-18, 1997; Burton et al., *Advances in Immunology*, 57:191-280, 1994; PCT application No. PCT/GB91/01134; PCT publications WO 90/02809; WO 91/10737; WO 92/01047; WO 92/18619; WO 93/11236; WO 95/15982; WO 95/20401; and
10 U.S. Patent Nos. 5,698,426; 5,223,409; 5,403,484; 5,580,717; 5,427,908; 5,750,753; 5,821,047; 5,571,698; 5,427,908; 5,516,637; 5,780,225; 5,658,727; 5,733,743 and 5,969,108; each of which is incorporated herein by reference in its entirety.

As described in the above references, after phage selection, the antibody coding regions from the phage can be isolated and used to generate whole antibodies, including
15 human antibodies, or any other desired fragments, and expressed in any desired host, including mammalian cells, insect cells, plant cells, yeast, and bacteria, e.g., as described in detail below. For example, techniques to recombinantly produce Fab, Fab' and F(ab')₂ fragments can also be employed using methods known in the art such as those disclosed in PCT publication WO 92/22324; Mullinax et al., *BioTechniques*, 12(6):864-869, 1992; and
20 Sawai et al., *AJRI*, 34:26-34, 1995; and Better et al., *Science*, 240:1041-1043, 1988 (each of which is incorporated by reference in its entirety). Examples of techniques which can be used to produce single-chain Fvs and antibodies include those described in U.S. Patent Nos. 4,946,778 and 5,258,498; Huston et al., *Methods in Enzymology*, 203:46-88, 1991; Shu et al., *PNAS*, 90:7995-7999, 1993; and Skerra et al., *Science*, 240:1038-1040, 1988.

- 25 Once an antibody molecule of the invention has been produced by any methods described above, it may then be purified by any method known in the art for purification of an immunoglobulin molecule, for example, by chromatography (e.g., ion exchange, affinity, particularly by affinity for the specific antigen after Protein A or Protein G purification, and sizing column chromatography), centrifugation, differential solubility, or by any other
30 standard techniques for the purification of proteins. Further, the antibodies of the present invention or fragments thereof may be fused to heterologous polypeptide sequences described herein or otherwise known in the art to facilitate purification.

For some uses, including *in vivo* use of antibodies in humans and *in vitro* detection assays, it may be preferable to use chimeric, humanized, or human antibodies. A chimeric antibody is a molecule in which different portions of the antibody are derived from different animal species, such as antibodies having a variable region derived from a murine

- 5 monoclonal antibody and a constant region derived from a human immunoglobulin. Methods for producing chimeric antibodies are known in the art. See e.g., Morrison, Science, 229:1202, 1985; Oi et al., BioTechniques, 4:214 1986; Gillies et al., J. Immunol. Methods, 125:191-202, 1989; U.S. Patent Nos. 5,807,715; 4,816,567; and 4,816,397, which are incorporated herein by reference in their entireties. Humanized antibodies are antibody
10 molecules from non-human species that bind the desired antigen having one or more complementarity determining regions (CDRs) from the non-human species and framework regions from a human immunoglobulin molecule. Often, framework residues in the human framework regions will be substituted with the corresponding residue from the CDR donor antibody to alter, preferably improve, antigen binding. These framework substitutions are
15 identified by methods well known in the art, e.g., by modeling of the interactions of the CDR and framework residues to identify framework residues important for antigen binding and sequence comparison to identify unusual framework residues at particular positions. See, e.g., Queen et al., U.S. Patent No. 5,585,089; Riechmann et al., Nature, 332:323, 1988, which are incorporated herein by reference in their entireties. Antibodies can be humanized
20 using a variety of techniques known in the art including, for example, CDR-grafting (EP 239,400; PCT publication WO 91/09967; U.S. Patent Nos. 5,225,539; 5,530,101 and 5,585,089), veneering or resurfacing (EP 592,106; EP 519,596; Padlan, Molecular Immunology, 28(4/5):489-498, 1991; Studnicka et al., Protein Engineering, 7(6):805-814, 1994; Roguska et al., Proc Natl. Acad. Sci. USA, 91:969-973, 1994), and chain shuffling
25 (U.S. Patent No. 5,565,332), all of which are hereby incorporated by reference in their entireties.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Human antibodies can be made by a variety of methods known in the art including phage display methods described above using antibody libraries derived from
30 human immunoglobulin sequences. See U.S. Patent Nos. 4,444,887 and 4,716,111; and PCT publications WO 98/46645; WO 98/50433; WO 98/24893; WO 98/16654; WO 96/34096; WO 96/33735; and WO 91/10741, each of which is incorporated herein by reference in its entirety.

Human antibodies can also be produced using transgenic mice which are incapable of expressing functional endogenous immunoglobulins, but which can express human immunoglobulin genes. For an overview of this technology for producing human antibodies, see Lonberg and Huszar, *Int. Rev. Immunol.*, 13:65-93, 1995. For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, e.g., PCT publications WO 98/24893; WO 92/01047; WO 96/34096; WO 96/33735; European Patent No. 0 598 877; U.S. Patent Nos. 5,413,923; 5,625,126; 5,633,425; 5,569,825; 5,661,016; 5,545,806; 5,814,318; 5,885,793; 5,916,771; and 5,939,598, which are incorporated by reference herein in their entireties. In addition, companies such as Abgenix, Inc. (Fremont, CA), Medarex (NJ) and Genpharm (San Jose, CA) can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, e.g., a mouse antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. (Jespers et al., *Bio/technology*, 12:899-903, 1988).

Antibodies fused or conjugated to heterologous polypeptides may be used in in vitro immunoassays and in purification methods (e.g., affinity chromatography) well known in the art. See e.g., PCT publication Number WO 93/21232; EP 439,095; Naramura et al., *Immunol. Lett.*, 39:91-99, 1994; U.S. Patent 5,474,981; Gillies et al., *PNAS*, 89:1428-1432, 1992; and Fell et al., *J. Immunol.*, 146:2446-2452, 1991, which are incorporated herein by reference in their entireties.

Antibodies may also be attached to solid supports, which are particularly useful for immunoassays or purification of the polypeptides of the invention or fragments, derivatives, analogs, or variants thereof, or similar molecules having the similar enzymatic activities as the polypeptide of the invention. Such solid supports include, but are not limited to, glass, cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or polypropylene.

30 5.6 Pharmaceutical Compositions and Kits

The present invention encompasses pharmaceutical compositions comprising anti-viral agents of the present invention. In a specific embodiment, the anti-viral agent is an

antibody which immunospecifically binds and neutralize the hSARS virus or variants thereof, or any proteins derived therefrom. In another specific embodiment, the anti-viral agent is a polypeptide or nucleic acid molecule of the invention. The pharmaceutical compositions have utility as an anti-viral prophylactic agent and may be administered to a subject where the subject has been exposed or is expected to be exposed to a virus.

Various delivery systems are known and can be used to administer the pharmaceutical composition of the invention, e.g., encapsulation in liposomes, microparticles, microcapsules, recombinant cells capable of expressing the mutant viruses, receptor mediated endocytosis (see, e.g., Wu and Wu, 1987, J. Biol. Chem. 262:4429 4432).

Methods of introduction include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, and oral routes. The compounds may be administered by any convenient route, for example by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (e.g., oral mucosa, rectal and intestinal mucosa, etc.) and may be administered together with other biologically active agents. Administration can be systemic or local. In a preferred embodiment, it may be desirable to introduce the pharmaceutical compositions of the invention into the lungs by any suitable route. Pulmonary administration can also be employed, e.g., by use of an inhaler or nebulizer, and formulation with an aerosolizing agent.

In a specific embodiment, it may be desirable to administer the pharmaceutical compositions of the invention locally to the area in need of treatment; this may be achieved by, for example, and not by way of limitation, local infusion during surgery, topical application, e.g., in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers. In one embodiment, administration can be by direct injection at the site (or former site) infected tissues.

In another embodiment, the pharmaceutical composition can be delivered in a vesicle, in particular a liposome (see Langer, 1990, Science 249:1527-1533; Treat et al., in Liposomes in the Therapy of Infectious Disease and Cancer, Lopez Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, ibid. , pp. 317-327; see generally ibid.).

In yet another embodiment, the pharmaceutical composition can be delivered in a controlled release system. In one embodiment, a pump may be used (see Langer, supra;

Sefton, 1987, CRC Crit. Ref. Biomed. Eng. 14:201; Buchwald et al., 1980, Surgery 88:507, and Saudek et al., 1989, N. Engl. J. Med. 321:574). In another embodiment, polymeric materials can be used (see Medical Applications of Controlled Release, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); Controlled Drug Bioavailability, Drug

- 5 Product Design and Performance, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, J. Macromol. Sci. Rev. Macromol. Chem. 23:61 (1983); see also Levy et al., 1985, Science 228:190; During et al., 1989, Ann. Neurol. 25:351; Howard et al., 1989, J. Neurosurg. 71:105). In yet another embodiment, a controlled release system can be placed in proximity of the composition's target, i.e., the lung, thus requiring only a fraction of the
10 systemic dose (see, e.g., Goodson, in Medical Applications of Controlled Release, supra, vol. 2, pp. 115-138 (1984)).

Other controlled release systems are discussed in the review by Langer (Science 249:1527-1533 (1990)).

- The pharmaceutical compositions of the present invention comprise a
15 therapeutically effective amount of an live attenuated, inactivated or killed hSARS virus, or recombinant or chimeric hSARS virus, and a pharmaceutically acceptable carrier. In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, and more particularly in
20 humans. The term "carrier" refers to a diluent, adjuvant, excipient, or vehicle with which the pharmaceutical composition is administered. Such pharmaceutical carriers can be sterile liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. Water is a preferred carrier when the pharmaceutical composition is administered intravenously.
25 Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid carriers, particularly for injectable solutions. Suitable pharmaceutical excipients include starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. The composition, if desired, can also contain minor amounts of
30 wetting or emulsifying agents, or pH buffering agents. These compositions can take the form of solutions, suspensions, emulsion, tablets, pills, capsules, powders, sustained release formulations and the like. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard

carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, etc. Examples of suitable pharmaceutical carriers are described in "Remington's Pharmaceutical Sciences" by E.W. Martin. The formulation should suit the mode of administration.

- 5 In a preferred embodiment, the composition is formulated in accordance with routine procedures as a pharmaceutical composition adapted for intravenous administration to human beings. Typically, compositions for intravenous administration are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the
10 injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the
15 composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

The pharmaceutical compositions of the invention can be formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and
20 those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2 ethylamino ethanol, histidine, procaine, etc.

The amount of the pharmaceutical composition of the invention which will be effective in the treatment of a particular disorder or condition will depend on the nature of
25 the disorder or condition, and can be determined by standard clinical techniques. In addition, *in vitro* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of the practitioner and each patient's circumstances. However,
30 suitable dosage ranges for intravenous administration are generally about 20 500 micrograms of active compound per kilogram body weight. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight to 1 mg/kg body

weight. Effective doses may be extrapolated from dose response curves derived from in vitro or animal model test systems.

Suppositories generally contain active ingredient in the range of 0.5% to 10% by weight; oral formulations preferably contain 10% to 95% active ingredient.

- 5 The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of
10 manufacture, use or sale for human administration. In a preferred embodiment, the kit contains an anti-viral agent of the invention, e.g., an antibody specific for the polypeptides encoded by a nucleotide sequence of SEQ ID NO:1, 11, 13, 15, 2471 or 2473, or as shown in Figures 11 (SEQ ID NOS:17-239, 241-736 and 738-1107) and 12 (SEQ ID NOS:1109-1589, 1591-1964 and 1966-2470), or any hSARS epitope, or a polypeptide or protein of the
15 present invention, or a nucleic acid molecule of the invention, alone or in combination with adjuvants, antivirals, antibiotics, analgesic, bronchodilators, or other pharmaceutically acceptable excipients.

The present invention further encompasses kits comprising a container containing a pharmaceutical composition of the present invention and instructions to for use.

20

5.7 Detection Assays

- The present invention provides a method for detecting an antibody, which immunospecifically binds to the hSARS virus, in a biological sample, for example blood, serum, plasma, saliva, urine, etc., from a patient suffering from SARS. In a specific
25 embodiment, the method comprising contacting the sample with the hSARS virus, for example, of deposit no. CCTCC-V200303, or having a genomic nucleic acid sequence of SEQ ID NO:15, directly immobilized on a substrate and detecting the virus-bound antibody directly or indirectly by a labeled heterologous anti-isotype antibody. In another specific embodiment, the sample is contacted with a host cell which is infected by the hSARS virus,
30 for example, of deposit no. CCTCC-V200303, or having a genomic nucleic acid sequence of SEQ ID NO:15, and the bound antibody can be detected by immunofluorescent assay as described in Section 6.5, *infra*.

An exemplary method for detecting the presence or absence of a polypeptide or nucleic acid of the invention in a biological sample involves obtaining a biological sample from various sources and contacting the sample with a compound or an agent capable of detecting an epitope or nucleic acid (e.g., mRNA, genomic RNA) of the hSARS virus such

- 5 that the presence of the hSARS virus is detected in the sample. A preferred agent for detecting hSARS mRNA or genomic RNA of the invention is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic RNA encoding a polypeptide of the invention. The nucleic acid probe can be, for example, a nucleic acid molecule comprising or consisting of the nucleotide sequence or SEQ ID NO:1, 11, 13, 15, 2471, or 2473, or a
10 portion thereof, such as an oligonucleotide of at least 15, 20, 25, 30, 50, 100, 250, 500, 750, 1,000 or more contiguous nucleotides in length and sufficient to specifically hybridize under stringent conditions to a hSARS mRNA or genomic RNA.

In another preferred specific embodiment, the presence of hSARS virus is detected in the sample by an reverse transcription polymerase chain reaction (RT-PCR) using the
15 primers that are constructed based on a partial nucleotide sequence of the genome of hSARS virus, for example, that of deposit accession no. CCTCC-V200303, or having a genomic nucleic acid sequence of SEQ ID NO:15, or based on a nucleotide sequence of SEQ ID NO:1, 11, 13, 2471 or 2473. In a non-limiting specific embodiment, preferred primers to be used in a RT-PCR method are: 5'-TACACACCTCAGC-GTTG-3' (SEQ ID
20 NO:3) and 5'-CACGAACGTGACG-AAT-3' (SEQ ID NO:4), in the presence of 2.5 mM MgCl₂ and the thermal cycles are, for example, but not limited to, 94 °C for 8 min followed by 40 cycles of 94 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min (*also see Sections 6.7 and 6.8 infra*). In preferred embodiments, the primers comprise nucleic acid sequence of SEQ ID NOS:2475 and 2476, or SEQ ID NOS:2480 and 2481. In preferred embodiments,
25 the thermal cycles are 94 °C for 10 min followed by 40 cycles of 94 °C for 30 seconds, 56 °C for 30 seconds, 72 °C for 30 seconds, 72°C for 10 minutes. In preferred embodiments, the primers comprise nucleic acid sequence of SEQ ID NOS:2477 and 2478. In more preferred specific embodiment, the present invention provides a real-time quantitative PCR assay to detect the presence of hSARS virus in a biological sample by subjecting the cDNA
30 obtained by reverse transcription of the extracted total RNA from the sample to PCR reactions using the specific primers, such as those having nucleotide sequences of SEQ ID NOS:3 and 4, and a fluorescence dye, such as SYBR® Green I, which fluoresces when bound non-specifically to double-stranded DNA. The fluorescence signals from these

reactions are captured at the end of extension steps as PCR product is generated over a range of the thermal cycles, thereby allowing the quantitative determination of the viral load in the sample based on an amplification plot (see Section 6.7, *infra*).

- A preferred agent for detecting hSARS is an antibody that specifically binds a polypeptide of the invention or any hSARS epitope, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')₂) can be used.

- The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The detection method of the invention can be used to detect mRNA, protein (or any epitope), or genomic RNA in a sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of mRNA include northern hybridizations, *in situ* hybridizations, RT-PCR, and RNase protection. *In vitro* techniques for detection of an epitope of hSARS include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of genomic RNA include nothern hybridizations, RT-PCT, and RNase protection. Furthermore, *in vivo* techniques for detection of hSARS include introducing into a subject organism a labeled antibody directed against the polypeptide. For example, the antibody can be labeled with a radioactive marker whose presence and location in the subject organism can be detected by standard imaging techniques, including autoradiography.
- In a specific embodiment, the methods further involve obtaining a control sample from a control subject, contacting the control sample with a compound or agent capable of detecting hSARS, e.g., a polypeptide of the invention or mRNA or genomic RNA encoding a polypeptide of the invention, such that the presence of hSARS or the polypeptide or mRNA or genomic RNA encoding the polypeptide is detected in the sample, and comparing the presence of hSARS or the polypeptide or mRNA or genomic RNA encoding the polypeptide in the control sample with the presence of hSARS, or the polypeptide or mRNA or genomic DNA encoding the polypeptide in the test sample.

The invention also encompasses kits for detecting the presence of hSARS or a polypeptide or nucleic acid of the invention in a test sample. The kit, for example, can comprise a labeled compound or agent capable of detecting hSARS or the polypeptide or a nucleic acid molecule encoding the polypeptide in a test sample and, in certain 5 embodiments, a means for determining the amount of the polypeptide or mRNA in the sample (e.g., an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits can also include instructions for use.

For antibody-based kits, the kit can comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide of the invention or hSARS epitope; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent. 10

For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic 15 acid sequence encoding a polypeptide of the invention or to a sequence within the hSARS genome or (2) a pair of primers useful for amplifying a nucleic acid molecule containing an hSARS sequence. The kit can also comprise, e.g., a buffering agent, a preservative, or a protein stabilizing agent. The kit can also comprise components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample 20 or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with instructions for use.

5.8 Screening Assays to Identify Anti-Viral Agents

25 The invention provides methods for the identification of a compound that inhibits the ability of hSARS virus to infect a host or a host cell. In certain embodiments, the invention provides methods for the identification of a compound that reduces the ability of hSARS virus to replicate in a host or a host cell. Any technique well-known to the skilled artisan can be used to screen for a compound that would abolish or reduce the ability of 30 hSARS virus to infect a host and/or to replicate in a host or a host cell.

In certain embodiments, the invention provides methods for the identification of a compound that inhibits the ability of hSARS virus to replicate in a mammal or a

mammalian cell. More specifically, the invention provides methods for the identification of a compound that inhibits the ability of hSARS virus to infect a mammal or a mammalian cell. In certain embodiments, the invention provides methods for the identification of a compound that inhibits the ability of hSARS virus to replicate in a mammalian cell. In a 5 specific embodiment, the mammalian cell is a human cell.

In another embodiment, a cell is contacted with a test compound and infected with the hSARS virus. In certain embodiments, a control culture is infected with the hSARS virus in the absence of a test compound. The cell can be contacted with a test compound before, concurrently with, or subsequent to the infection with the hSARS virus. In a 10 specific embodiment, the cell is a mammalian cell. In an even more specific embodiment, the cell is a human cell. In certain embodiments, the cell is incubated with the test compound for at least 1 minute, at least 5 minutes at least 15 minutes, at least 30 minutes, at least 1 hour, at least 2 hours, at least 5 hours, at least 12 hours, or at least 1 day. The titer of the virus can be measured at any time during the assay. In certain embodiments, a time 15 course of viral growth in the culture is determined. If the viral growth is inhibited or reduced in the presence of the test compound, the test compound is identified as being effective in inhibiting or reducing the growth or infection of the hSARS virus. In a specific embodiment, the compound that inhibits or reduces the growth of the hSARS virus is tested for its ability to inhibit or reduce the growth rate of other viruses to test its specificity for the 20 hSARS virus.

In one embodiment, a test compound is administered to a model animal and the model animal is infected with the hSARS virus. In certain embodiments, a control model animal is infected with the hSARS virus without the administration of a test compound. The test compound can be administered before, concurrently with, or subsequent to the 25 infection with the hSARS virus. In a specific embodiment, the model animal is a mammal. In an even more specific embodiment, the model animal can be, but is not limited to, a cotton rat, a mouse, or a monkey. The titer of the virus in the model animal can be measured at any time during the assay. In certain embodiments, a time course of viral growth in the culture is determined. If the viral growth is inhibited or reduced in the 30 presence of the test compound, the test compound is identified as being effective in inhibiting or reducing the growth or infection of the hSARS virus. In a specific embodiment, the compound that inhibits or reduces the growth of the hSARS in the model

animal is tested for its ability to inhibit or reduce the growth rate of other viruses to test its specificity for the hSARS virus.

6. EXAMPLES

- 5 The following examples illustrate the isolation and identification of the novel hSARS virus. These examples should not be construed as limiting.

METHODS AND RESULTS

- As a general reference, Wiedbrauk DL & Johnston SLG. (Manual of Clinical
10 Virology, Raven Press, New York, 1993) was used.

6.1 Clinical Subjects

- The study included all 50 patients who fitted a modified World Health Organization (WHO) definition of SARS and were admitted to 2 acute regional hospitals in Hong Kong
15 Special Administrative Region (HKSAR) between February 26 to March 26, 2003 (WHO. Severe acute respiratory syndrome (SARS) *Weekly Epidemiol Rec.* 2003; **78:** 81-83). A lung biopsy from an additional patient, who had typical SARS and was admitted to a third hospital, was also included in the study. Briefly, the case definition for SARS was: (i) fever of 38°C or more; (ii) cough or shortness of breath; (iii) new pulmonary infiltrates on chest
20 radiograph; and (iv) either a history of exposure to a patient with SARS or absence of response to empirical antimicrobial coverage for typical and atypical pneumonia (beta-lactams and macrolides, fluoroquinolones or tetracyclines).

- Nasopharyngeal aspirates and serum samples were collected from all patients.
Paired acute and convalescent sera and feces were available from some patients. Lung
25 biopsy tissue from one patient was processed for a viral culture, RT-PCR, routine histopathological examination, and electron microscopy. Nasopharyngeal aspirates, feces and sera submitted for microbiological investigation of other diseases were included in the study under blinding and served as controls.

- The medical records were reviewed retrospectively by the attending physicians and
30 clinical microbiologists. Routine hematological, biochemical and microbiological examinations, including bacterial culture of blood and sputum, serological study and collection of nasopharyngeal aspirates for virological tests, were carried out.

6.2 Cell Line

FRhK-4 (fetal rhesus monkey kidney) cells were maintained in minimal essential medium (MEM) with 1% fetal calf serum, 1% streptomycin and penicillin, 0.2% nystatin
5 and 0.05% garamycin.

6.3 Viral Infection

Two-hundred µl of clinical (nasopharyngeal aspirates) samples, from two patients (see the Result section, *infra*), in virus transport medium were used to infect FRhK-4 cells.
10 The inoculated cells were incubated at 37°C for 1 hour. One ml of MEM containing 1 µg trypsin was then added to the culture and the infected cells were incubated in a 37°C incubator supplied with 5% carbon dioxide. Cytopathic effects were observed in the infected cells after 2 to 4 days of incubation. The infected cells were passaged into new FRhK-4 cells and cytopathic effects were observed within 1 day after the inoculation. The
15 infected cells were tested by an immunofluorescent assay for influenza A, influenza B, respiratory syncytial virus, parainfluenza types 1, 2 and 3, adenovirus and human metapneumovirus (hMPV) and negative results were obtained for all cases. The infected cells were also tested by RT-PCR for influenza A and human metapneumovirus with negative results.
20

6.4 Virus Morphology

The infected cells prepared as described above were harvested, pelleted by centrifugation and the cell pellets were processed for thin-section transmitted electron microscopic visualization. Viral particles were identified in the cells infected with both
25 clinical specimens, but not in control cells which were not infected with the virus. Virions isolated from the infected cells were about 70-100 nanometers (Figure 2). Viral capsids were found predominantly within the vesicles of the golgi and endoplasmic reticulum and were not free in the cytoplasm. Virus particles were also found at the cell membrane.

One virus isolate was ultracentrifuged and the cell pellet was negatively stained
30 using phosphotungstic acid. Virus particles characteristic of *Coronaviridae* were thus visualized. Since the human *Coronaviruses* hitherto recognized are not known to cause a

similar disease, the present inventors postulated that the virus isolates represent a novel virus that infects humans.

6.5 Antibody Response to the Isolated Virus

5 To further confirm that this novel virus is responsible for causing SARS in the infected patients, blood serum samples from the patients who were suffering from SARS were obtained and a neutralization test was performed. Typically diluted serum ($\times 50$, $\times 200$, $\times 800$ and $\times 1600$) was incubated with acetone-fixed FRhK-4 cells infected with hSARS at 37°C for 45 minutes. The incubated cells were then washed with phosphate-buffered saline
10 and stained with anti-human IgG-FITC conjugated antibody. The cells were then washed and examined under a fluorescent microscope. In these experiments, positive signals were found in 8 patients who had SARS (Figure 3), indicating that these patients had an IgG antibody response to this novel human respiratory virus of *Coronaviridae*. By contrast, no signal was detected in 4 negative-control paired sera. The serum titers of anti- hSARS
15 antibodies of the tested patients are shown in Table 1.

Table 1

Name	Date	Lab No.	Anti-SARS
Patient A	25-Feb-03	S2728	<50
	6-Mar-03	S2728	1600
Patient B	26-Feb-03	S2441	50
	3-Mar-03	S2441	200
Patient C	4-Mar-03	S3279	200
	14-Mar-03	S3279	1600
Patient D	6-Mar-03	M41045	<50
	11-Mar-03	MB943703	800
Patient E	4-Mar-03	M38953	<50
	18-Mar-03	KWH03/3601	800
Control F	13-Feb-03	M27124	<50
	1-Mar-03	MB942968	<50
Patient G	3-Mar-03	M38685	<50
	7-Mar-03	KWH03/2900	Equivocal

Blinded samples:

1a *	Acute	<50
1b	Convalescent	1600
2a *	Acute	50
2b	Convalescent	>1600
3a *	Acute	50
3b	Convalescent	>1600
4a *	Acute	<50
4b	Convalescent	<50
5a *	Acute	<50
5b	Convalescent	<50
6a *	Acute	<50
6b	Convalescent	<50

NB: * patients with SARS

These results indicated that this novel member of *Coronaviridae* is a key pathogen
5 in SARS.

6.6 Sequences of the hSARS Virus

Total RNA from infected or uninfected FrHK-4 cells was harvested two days post-infection. One-hundred ng of purified RNA was reverse transcribed using Superscript II
10 reverse transcriptase (Invitrogen) in a 20 µl reaction mixture containing 10 pg of a degenerated primer (5'-GCCGGAGCTCTGCAGAATTCCNNNNNN-3', N=A, T, G or C: SEQ ID NO:5) as recommended by the manufacturer. Reverse transcribed products were then purified by a QIAquick PCR purification kit as instructed by the manufacturer and

eluted in 30 µl of 10 mM Tris-HCl, pH 8.0 . Three µl of purified cDNA products were add in a 25 µl reaction mixture containing 2.5 µl of 10x PCR buffer, 4 µl of 25mM MgCl₂, 0.5 µl of 10 mM dNTP, 0.25 µl of AmpliTaq Gold® DNA polymerase (Applied Biosystems), 2.5 µCi of [α -³²P]CTP (Amersham), 2 µl of 10 µM primer (5'

- 5 GCCGGAGCTCTGCAGAATT-C-3': SEQ ID NO:6). Reactions were thermal cycled through the following profile: 94°C for 8 min followed by 2 cycles of 94°C for 1 min, 40°C for 1 min, 72°C for 2 min. This temperature profile was followed by 35 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 1 min. 6 µl of the PCR products were analyzed in a 5% denaturing polyacrylamide gel electrophoresis. Gel was exposed to X-ray film and the
10 film was developed after an over-night exposure. Unique PCR products which were only identified in infected cell samples were isolated from the gel and eluted in a 50 µl of 1x TE buffer. Eluted PCR products were then re-amplified in 25 µl of reaction mixture containing 2.5 µl of 10x PCR buffer, 4 µl of 25 mM MgCl₂, 0.5 µl ru 10 mM dNTP, 0.25 µl of AmpliTaq Gold® DNA polymerase (Applied Biosystems), 1 µl of 10 µM primer (5'-
15 GCCGGAGCTCTGCAGAATTC-3':SEQ ID NO:6). Reaction mixtures were thermal cycled through the following profile: 94°C for 8 min followed by 35 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 1 min. PCR products were cloned using a TOPO TA cloning kit (Invitrogen) and ligated plasmids were transformed into TOP10 *E. coli* competent cells (Invitrogen). PCR inserts were sequenced by a BigDye cycle sequencing kit as
20 recommended by the manufacturer (Applied Biosystems) and sequencing products were analyzed by an automatic sequencer (Applied Biosystems, model number 3770). The obtained sequence (SEQ ID NO:1) is shown in Figure 1. The deduced amino acid sequence (SEQ ID NO:2) from the obtained DNA sequence showed 57% homology to the polymerase protein of identified *coronaviruses*.

25 Similarly, two other partial sequences (SEQ ID NOS:11 and 13) and deduced amino acid sequences (SEQ ID NOS:12 and 14, respectively) were obtained from the hSARS virus and are shown in Figures 8 (SEQ ID NOS:11 and 12) and 9 (SEQ ID NOS:13 and 14).

The entire genomic sequence of hSARS virus is shown in Figure 10 (SEQ ID NO:15). The deduced amino acid sequences of SEQ ID NO:15 in all three frames (SEQ ID
30 NO:16, 240 and 737) are shown in Figure 11 (SEQ ID NOS:17-239, 241-736 and 738-1107). The deduced amino acid sequences of the complement of SEQ ID NO:15 in all three

frames (SEQ ID NOS:1108, 1590 and 1965) are shown in Figure 12 (SEQ ID NOS:1109-1589, 1591-1964 and 1966-2470).

6.7 Detection of hSARS Virus in Nasopharyngeal Aspirates

5 First, the nasopharyngeal aspirates (NPA) were examined by rapid immunofluorescent antigen detection for influenza A and B, parainfluenza types 1, 2 and 3, respiratory syncytial virus and adenovirus (Chan KH, Maldeis N, Pope W, Yup A, Ozinskas A, Gill J, Seto WH, Shortridge KF, Peiris JSM. Evaluation of Directigen Fly A+B test for rapid diagnosis of influenza A and B virus infections. *J Clin Microbiol.* 2002; **40:** 1675-1680) and were cultured for conventional respiratory pathogens on Mardin Darby Canine Kidney, LLC-Mk2, RDE, Hep-2 and MRC-5 cells (Wiedbrauk DL, Johnston SLG. *Manual of clinical virology*. Raven Press, New York. 1993). Subsequently, fetal rhesus kidney (FRhk-4) and A-549 cells were added to the panel of cell lines used. Reverse transcription polymerase chain reaction (RT-PCR) was performed directly on the clinical specimen for 10 influenza A (Fouchier RA, Bestebroer TM, Herfst S, Van Der Kemp L, Rimmelzwaan GF, Osterhaus AD. Detection of influenza A virus from different species by PCR amplification of conserved sequences in the matrix gene. *J Clin Microbiol.* 2000; **38:** 4096-101) and 15 human metapneumovirus (HMPV). The primers used for HMPV were: for first round, 5'-AARGTSAATGCATCAGC-3' (SEQ ID NO: 7) and 5'-CAKATTYTGCTTATGCTTC-3' (SEQ ID NO:8); and nested primers: 5'-ACACCTGTTACAATACCAGC-3' (SEQ ID NO:9) and 5'-GACTTGAGTCCCAGCTCA-3' (SEQ ID NO:10). The size of the nested 20 PCR product was 201 bp. An ELISA for mycoplasma was used to screen cell cultures (Roche Diagnostics GmbH, Roche, Indianapolis, USA).

25 **RT-PCR Assay**

Subsequent to culturing and genetic sequencing of the hSARS virus from two patients (see Section 6.6, *supra*), an RT-PCR was developed to detect the hSARS virus sequence from NPA samples. Total RNA from clinical samples was reverse transcribed using random hexamers and cDNA was amplified using primers 5'-TACACACCTCAGC-30 GTTG-3' (SEQ ID NO:3) and 5'-CACGAACGTGACGAAT-3' (SEQ ID NO:4), which are constructed based on the in the presence of 2.5 mM MgCl₂ (94 °C for 8 min followed by 40 cycles of 94 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min).

The summary of a typical RT-PCR protocol is as follows:

1. RNA extraction

RNA from 140 µl of NPA samples is extracted by QIAquick viral RNA extraction
5 kit and is eluted in 50 µl of elution buffer.

2. Reverse transcription

	RNA	11.5 µl
	0.1 M DTT	2 µl
10	5x buffer	4 µl
	10 mM dNTP	1 µl
	Superscript II, 200 U/µl (Invitrogen)	1 µl
	Random hexamers, 0.3 µg/ µl	0.5 µl
	Reaction condition	42 °C, 50 min
		94 °C, 3 min
		4 °C

3. PCR

20 cDNA generated by random primers is amplified in a 50 µl reaction as follows:

	cDNA	2 µl
	10 mM dNTP	0.5 µl
	10x buffer	5 µl
25	25 mM MgCl ₂	5 µl
	25 µM Forward primer	0.5 µl
	25 µM Reverse primer	0.5 µl
	AmpliTaq Gold polymerase, 5U/µl (Applied Biosystems)	0.25 µl
	Water	36.25 µl

30

Thermal-cycle condition: 95°C, 10 min, followed by 40 cycles of 95 °C, 1 min; 50°C 1 min; 72 °C, 1 min.

4. Primer sequences

5 Primers were designed based on the RNA-dependent RNA polymerase encoding sequence (SEQ ID NO:1) of the hSARS virus.

Forward primer: 5' TACACACCTCAGCGTTG 3' (SEQ ID NO:3)

Reverse primer: 5' CACGAACGTGACGAAT 3' (SEQ ID NO:4)

10 Product size: 182 bps

Real-Time Quantitative PCR Assay

Total RNA from 140 µl of nasopharyngeal aspirate (NPA) was extracted by QIAamp® virus RNA mini kit (Qiagen) as instructed by the manufacturer. Ten µl of eluted 15 RNA samples were reverse transcribed by 200 U of Superscript II® reverse transcriptase (Invitrogen) in a 20 µl reaction mixture containing 0.15 µg of random hexamers, 10 mmol/l DTT, and 0.5 mmol/l dNTP, as instructed. Complementary DNA was then amplified in a SYBR® Green I fluorescence reaction (Roche) mixtures. Briefly, 20 µl reaction mixtures containing 2 µl of cDNA, 3.5 mmol/l MgCl₂, 0.25 µmol/l of forward primer (5'- 20 TACACACCTCAGCGTTG-3'; SEQ ID NO:3) and 0.25 µmol/l reverse primer (5'- CACGAACGTGACGAAT-3'; SEQ ID NO:4) were thermal-cycled by a Light-Cycler (Roche) with the PCR program, [95°C, 10 min followed by 50 cycles of 95°C, 10 min; 57°C, 5 sec; 72°C 9 sec]. Plasmids containing the target sequence were used as positive controls. Fluorescence signals from these reactions were captured at the end of extension 25 step in each cycle. To determine the specificity of the assay, PCR products (184 base pairs) were subjected to a melting curve analysis at the end of the assay (65°C to 95°C, 0.1 °C per second).

6.8 Detection of N-gene of hSARS Virus in Patients

30 **6.8.1 RT-PCR diagnosis protocol for coronavirus in SARS patients**

Equipment required (for 96 samples):

1 × SV Total RNA Isolation system
2 × Mega titer plate
3 × 96-well PCR plate
1 × 0.5-10 µl multi-channel pipette
5 1 × 10-100 µl multi-channel pipette
1 × 20-200 µl multi-channel pipette
1 × vacuum pump
1 × swing-bucket rotor with microtest plate buckets
2 × PCR machine (96-well plate compatible)
10 1 × Gel electrophoresis apparatus

Station 1* – clinical samples handling (1 medical officer/ clinical technician)

- Aliquot 500 µl sample in viral transport medium (containing, per liter, 2g of sodium bicarbonate, 5 g of bovine serum albumin, 200 µg of vancomycin, 18 µg of amikacin, and 160 U of nystatin in Earle's balanced salt solution) from each individual vial into a well of 96-well mega titer plate containing 500 µl lysis buffer (1x) containing 100 µl PK-15 cell (ATCC CCL-33; 5.0×10^5 cell/ml) in complete minimum essential medium with Earle's salt (EMEM, Invitrogen) as internal control.**
- Mix the lysate by pipetting up-and-down 3 times
- Proceed to Station 2.

*Station 1 should be carried out inside Class III biological safety cabinet.

** At least two negative samples should be included in a 96-well platform as a negative control.

25 **Station 2 – Total RNA extraction (1 laboratory technician)**

- Set up the Vacuum Manifold unit. Place the binding plate onto the Manifold Base.
- Transfer the lysate from mega titer plate to each well of the SV 96 Binding Plate (binding plate).
- 30 • Apply vacuum until the lysate passes through the binding plate. Release vacuum.
- Add 500 µl of SV RNA Wash Solution (wash solution) to each well of binding plate.
- Apply vacuum until the wash solution passes through binding plate. Release vacuum.

- Prepare DNase incubation mix for an entire 96-well plate as below:

Yellow Core Buffer	2 ml
0.09 M MnCl ₂	250 µl
DNase I	250 µl

- 5 ○ Apply 25 µl freshly prepared DNase incubation mix directly to the membrane of the binding plate.
- Incubate at 20-25°C for 10 minutes.
- Add 200 µl of SV DNase Stop Solution to each well of the binding plate.
- Apply vacuum until the SV DNase Stop Solution passes through the binding plate.
- 10 Release vacuum.
- Add 500 µl wash solution to each well of the binding plate.
- Apply vacuum until wash solution passes through the binding plate. Turn off vacuum.
- Spin the binding plate at 3000 ×g for 30 seconds to remove residue wash solution.
- 15 ● Transfer the binding plate on top of a 96-well RT plate.
- Add 50 µl nuclease-free water into each well of the binding plate to elute RNA.
- Incubate at room temperature for 1 minute.
- Spin the binding plate at 3000 ×g, 4°C for 1 minute.
- Collect eluted RNA in the 96-well RT plate.
- 20 ● Add 5 µl of 3 M sodium acetate and 200 µl of 95% ethanol into each well of the plate.
- Place the RT plate on ice and incubate for 15 minutes.
- Spin the plate at 3000 ×g, 4°C, 15 minutes.
- Discard supernatant by inverting the plate and blotting on a clean paper towel.
- 25 ● Wash the pellet with 200µl of 70% ethanol.
- Spin the plate at 3000×g, 4°C, 10 minutes.
- Discard supernatant by inverting the plate and blotting on a clean paper towel.
- Air-dry the pellet for 5 minutes.
- Add 12 µl of nuclease-free water into each well.
- 30 ● Vortex the plate briefly to dissolve the pellet (for an example result, see Fig. 18).
- Proceed to Station 3.

Station 3 – Reverse transcription (1 laboratory technician)

- Prepare RT master mix for an entire 96-well plate in a 1.5-ml tube as below (100 reactions):

5

		<u>Per Reaction</u>	<u>×100</u>
	□ Random hexamers, 3µg/µl	0.05 µl	5 µl
	■ DNTPs, 10 mM	1 µl	100 µl
	□ First-strand buffer, 5×	4 µl	400 µl
10	□ DTT, 0.1 M	2 µl	200 µl
	■ Superscript II, 200U/µl	1 µl	100 µl
	Total	8.05 µl	805 µl

- Aliquot 100 µl RT mix into 8 wells of a clean 96-well master mix plate.
- From this plate, transfer 8.05 µl RT mix to each well of RT plate containing 12 µl RNA, mix by pipetting up-and-down for 3 times with a multi channel pipette.
REPLACE TIP AFTER EACH TRANSFER.
- Incubate the samples at 42°C for 50minutes followed by 70°C for 15 minutes.
- Proceed to Station 4.

20

Station 4 – N-gene specific PCR (1 laboratory technician)

- Prepare PCR master mix for an entire 96-well plate in two 2059 culture tubes as below (100 reactions):

	<u>N-specific PCR</u>		<u>Control PCR</u>		
	Per 25 µl Reaction	×100	Per 25 µl Reaction	×100	
25	mQH ₂ O	18.65 µl	1865 µl	17.65 µl	1765 µl
	10×PCR buffer	2.5 µl	250 µl	2.5 µl	250 µl
	25 mM MgCl ₂	1.5 µl	150 µl	2.5 µl	250 µl
	10 mM dNTPs	0.25 µl	25 µl	0.25 µl	25 µl
30	Forward primer 10 µM	0.5 µl	50 µl	0.5 µl	50 µl
	Reverse primer 10 µM	0.5 µl	50 µl	0.5 µl	50 µl
	AmpliTaq Gold® 500 U	0.1 µl	10 µl	0.1 µl	10 µl

Template DNA	1 µl	-	1 µl	-
Total	25 µl	2400 µl	25 µl	2400 µl

- N-gene specific PCR and control PCR are performed in two individual PCR plates.
- 5
 - Aliquot 290 µl PCR master mix into the first column of a 96-well PCR plate.
 - From the first column, aliquot 24 µl of master mix into each well of PCR plate.
 - Transfer 1 µl of cDNA template (from station 4) into each well of PCR plate.
 - Mix by pipetting up-and-down for 3 times with a multi-channel pipette. REPLACE TIP AFTER EACH TRANSFER.
- 10
 - Seal the plate with sealing tape.
 - Perform the following reaction in two 96-well PCR machines:

	<u>N-gene specific PCR</u>	<u>Control PCR</u>
	94°C 10 minutes	94°C 10 minutes
15	94°C 30 seconds } 56°C 30 seconds } 40 cycles	94°C 30 seconds } 55°C 30 seconds } 35 cycles
	72°C 30 seconds	72°C 45 seconds
	72°C 10 minutes	72°C 10 minutes

- 20 Station 5 – Gel electrophoresis (1 laboratory technician)
- Mix 5µl of N-gene specific PCR product and 5µl control PCR product with 1µl bromophenol blue loading dye
 - Load the samples into the wells of a 2% agarose gel.
 - Electrophoresize the PCR products at 140V, 250mA for 30 minutes.
 - 25 • Stain the gel with ethidium bromide.
 - Visualize the products with UV and record the result.

6.8.2 Using primers of SEQ ID NOS:2480 and 2481

RT-PCR diagnostic protocol was performed as described in Section 6.8.1 with some
30 modifications.

RNA isolation from clinical samples

Clinical samples including nasopharyngeal aspirates (NPA) and stool specimens were provided by the Department of Microbiology, The University of Hong Kong. In addition, tracheal dispersion and lung biopsy from an index patient A described in New Engl. J. Med. 348:1967-76 (by Drosten C.S., *et al.*, 2003) at three time points was also collected. Sample collection was conducted from April 1 to April 28, 2003 in local hospitals. Method of sample collection was described in the previous section (*also see*, Poon *et al.*, 2003, *Clinical Chemistry* 49:953-955). Total RNA extraction from clinical samples was carried out with SV96 Total RNA Isolation System (Promega, WI, USA), with following modifications from manufacturer's protocol. Five-hundred (500) µl of NPA/stool sample in viral transport medium (containing, per liter, 2 g of sodium bicarbonate, 5 g of bovine serum albumin, 200 µg of vancomycin, 18 µg of amikacin, and 160 U of nystatin in Earle's balanced salt solution) was mixed with equal volume of SV RNA Lysis Buffer containing 100 µl of pig kidney epithelial (PK-15) cell (ATCC CCL-33; 5.0x10⁵ cells/ml) in complete minimum essential medium with Earle's salt (EMEM, Invitrogen) as internal control. The mixture was transferred to the wells of the SV 96 Binding Plate. After washing with 500 µl of SV RNA Wash Solution prior to elution step, the plate was spun at 3000 xg for 30 seconds to remove residue wash solution. RNA was then eluted with 50 µl of nuclease-free water, and was collected in a clean 96-well PCR plate by spinning the plate at 3000 xg for 1 minute. Eluted RNA was then concentrated by incubating on ice for 15 minutes, in the presence of 5 µl of 3 M sodium acetate and 200 µl of 95 % ethanol. After centrifugation at 3000 xg, 4°C for 15 minutes, RNA pellet was washed with 200 µl of 75% ethanol and dissolved with 12 µl of nuclease-free water. Extracted RNA was immediately reverse-transcribed to first-strand cDNA.

25

First-Strand cDNA Synthesis

Reverse-transcription was performed with 200 U of Superscript® II reverse transcriptase (Invitrogen, USA) in a 20 µl reaction containing 0.15 µg of random hexamers, RT buffer (1x), 10 mM dithiothreitol (DTT) and 0.5 mM deoxynucleotide triphosphates (dNTPs). Reaction was carried out in Peltier Thermal Cycler (MJ Research) with the following conditions: 50 minutes at 42°C followed by 15 minutes at 70°C.

Polymerase Chain Reaction (PCR)

Primers were designed according to complete SARS CoV genomic sequence of a local specimen HK-39 announced previously (accession no. AY278491). Forward primer (SRS251: 5'-GCAGTCAAGCCTTCTCG-3'; SEQ ID NO:2480, corresponding to nt 5 28658-28676 of HK-39 SARS genome, i.e., CCTCC200303) and reverse primer (SRS252: 5'-GCCTCAGCAGCAGATTTC-3', SEQ ID NO:2481; corresponding to nt 28866-28883 of HK-39 SARS genome) amplified a 225 bp fragment from the region of N-gene that showed no homology to other coronavirus. Primers amplifying RNA-dependent RNA polymerase (1b gene) were used as parallel control (coro3: 5'-TACACACCTCAGCGTTG-10 3' (SEQ ID NO:3), corresponding to nt 18041-18057; and coro4: 5'-CACGAACGTGACGAAT-3' (SEQ ID NO:4), corresponding to nt 18207-18222, Department of Microbiology, the University of Hong Kong). Both amplicons were cloned into same pCR2.1 cloning vector (Fig. 17). Serially diluted plasmid was then used to determine the dynamic range and optimal condition of the PCRs (Fig. 21A and 21B).

15 Another set of primer that amplifying a 745 bp fragment from pig β -actin gene was employed as an internal control for the diagnostic PCR assay (actin-F: 5' - TGAGACCTAACACGCC-3' (SEQ ID NO:2482); and actin-R: 5' - ATCTGCTGAAAGGTGGAC-3' (SEQ ID NO:2483)).

Conventional PCR and gel electrophoresis was carried out as preliminary experiment.

20 Briefly, 1 μ l of cDNA from clinical samples was amplified with 0.5 U Taq DNA polymerase recombinant (Invitrogen Life Technologies) in a 25- μ l reaction containing PCR buffer (1x), 1.5 mM MgCl₂, 0.1 mM dNTPs and 0.5 pmol of each forward and reverse primers. Reaction was performed in Peltier Thermal Cycler (MJ Research) with the following conditions: 3 minutes at 94 °C, followed by 50 cycles of 94°C for 10 seconds, 56 °C for 10 seconds, 72 °C for 10 seconds, and a 10-minute final extension step at 72 °C.

25 Amplicons were analyzed with 2 % agarose gel electrophoresis (Fig. 23). Quantitative real-time PCR using SYBR® SYBR® green fluorophore was performed in diagnosis of clinical samples. In a 25 μ l reaction, 1 μ l cDNA template was mixed with 12.5 μ l (2x) Green PCR Master Mix (Applied Biosystems) and 0.5 pmol of each forward and reverse primer.

30 Volume of the reaction was adjusted to 25 μ l with distilled water. Reactions were performed in the iCycler iQ Real-Time PCR Detection System (Bio-Rad) under the same condition as the conventional PCR. Fluorescence signals (FAM, excitation = 490 nm,

emission = 530 nm) were collected at the end of each extension step during the PCR cycles (Fig. 22A). Threshold cycle (C_t) of each sample was determined using maximum curvature approach. Melting curve analysis was performed after 10 minutes final extension (Fig. 22B). cDNA from non-SARS patients, including patients suffering from adenovirus (n = 5), 5 respiratory syncytial virus (n = 5), human metapneumovirus (n = 5), influenza A virus (n = 5), or influenza B virus (n = 5) infection, were used as negative controls for the assay.

Northern Blot Analysis

SARS-CoV HK-39 strain infected Vero cell was provided by Department of Microbiology, the University of Hong Kong. Total RNA was extracted from the cell with TRIzol[®] reagent (Invitrogen Life Technologies) according to the manufacturer's protocol. Eight (8) µg of total RNA was separated by electrophoresis on a 1% agarose gel containing 3.7% formaldehyde. RNA was transferred to a positively charged nylon membrane Roche Diagnostic Corporation) by capillary blotting and fixed by UV cross-linking. cDNA 10 synthesized with the same RNA sample was used as template for probe synthesis. Four pairs of primers amplifying fragments from 1b (nt 18057 – 18222; SEQ ID NO:2484), S (nt 21920 – 22107; SEQ ID NO:2485), M (nt 26867 – 26996; SEQ ID NO:2486) and N (nt 28658 – 28883; SEQ ID NO:2487) gene were used in probe synthesis. DIG-labeling of probes, hybridization and detection of bands were performed with the digoxigenin system 15 according to the manufacturer's procedures (Roche Molecular Biochemicals). Signals were then analysed with chemiluminescence (Fig. 24).

Results and Discussion

A large-scale RT-PCR assay provides a rapid means in monitoring and screening of 20 SARS suspects. The result can be used to complement clinical diagnostic evaluation. In order to achieve a diagnostic purpose, the assay should be reliable and its accuracy should be assured so as to prevent occurrence of both false negative and false positive results. However, accuracy of the test may be influenced by several factors. A common technical problem with PCR is a failure of amplification due to the presence of PCR inhibitors (see 25 Fig. 21).

These PCR inhibitors included heme compounds found in blood, aqueous and vitreous humors, heparin, EDTA, urine and polyamines (Fredricks *et al.*, 1998, *J. Clin.*

Micro. 36:2810-16). Currently, NPA or stool samples were collected into transport medium to maintain the viability of the viral particles. RT-PCR was inhibited when total RNA extracted was used directly for first-strand cDNA synthesis without any treatment (25 out of 27 samples) in preliminary experiment. However, after a simple ethanol precipitation step, the amplification of DNA could be retained (Fig. 19). Same result was obtained by either using SV or SV96 total RNA Isolation System (data not shown). It demonstrated that some components either in the medium or NPA/ stool samples would affect the downstream processes of the diagnosis test.

In addition, current sample collection procedure dilutes the virus titer in the samples, especially during early stage of infection, in which the virus titer is low in nasal and throat swab specimens (Drosten *et al.*, 2003, *New England Journal of Medicine*, on-line at <http://content.nejm.org/cgi/reprint/NEJMoa030747v2>). It was suggested that the sensitivity of PCR tests for SARS depended on the quality of the specimen and the time of testing during the course of the illness. In order to increase sensitivity of the test, total RNA isolated from clinical samples was concentrated prior to 1st strand cDNA synthesis.

In order to avoid false negative PCR results due to failure in the process of RNA isolation and 1st strand cDNA synthesis, total RNA was extracted from clinical samples in parallel with PK-15 mammalian cells. Figure 23 showed the RT-PCR screening result on 48 clinical samples, including both NPA and stool samples. Diagnostic PCR was performed in parallel with β -actin PCR. All samples were positive in β -actin PCR. The result indicated that RNA and cDNA could be extracted and synthesized successfully from the samples in a single-step protocol as disclosed herein. With this internal control, total RNA isolation and cDNA synthesis from the samples were ensured, which eliminated false negative that resulted from failure in either one of the above processes. Moreover, 96-well assay format currently developed can be adopted into a high-throughput screening protocol, with which we are able to obtain diagnostic result of more than 90 clinical samples in 3 hours with 1 clinical personnel, while the current existing protocol, in which samples are proceeded in individual tubes, can only handle about 30-50 samples a day per technician.

Real-time quantitative PCR assay is more sensitive than conventional agarose gel-electrophoresis-associated PCR assay (Poon *et al.*, *J. Clin. Virol.* 28:233-8) and therefore employed for SARS-CoV diagnosis purpose. Positive signals were detected in 38 of 136 randomly selected clinical samples in both N-gene and 1b-gene specific PCR. Among these 38 positives, 3 were stool samples (2.2%) and 35 were NPA samples (25.7%). Detection

rate of the assay employing N-gene specific RT-PCR at different time points was shown in Table 2.

Table 2

Date of onset	No. of sample	No. of positive	Detection rate (%)
1-2	15	2	13.3
3-4	17	4	23.5
5-6	15	4	26.7
7-8	13	5	38.5
9-10	9	4	44.4
Negative control	19	All negative	--

5

10

15 Affirmative of these 38 positive cases was confirmed by melting curve analysis of PCR products. Specific melting temperature of N gene and 1b gene PCR products (85.5°C and 80.5°C, respectively) indicated that the target fragments were amplified in the reaction. Specificity of the assay was also validated with non-SARS patients samples, including patients suffering from adenovirus (n = 5), respiratory syncytial virus (n = 5), human 20 metapneumovirus (n = 5), influenza A virus (n = 5) and influenza B virus (n = 5). The result shows that all of these samples were negative in the assay (Fig. ??). These results indicate that the N-gene specific RT-PCR assay is specific for SARS-CoV diagnosis.

Furthermore, we also demonstrated that the N-gene specific PCR was more sensitive than that of PCR amplifying 1b RNA polymerase gene. Amplification conditions for both 25 PCR assays were optimized (see Fig. 22) first with the plasmid construct containing 1:1 ratio of 1b- and N-gene fragment (see Fig. 20). Dynamic range of N-gene specific PCR was obtained (Fig. ??) and it was found to be with lower Ct values than that of 1b-specific PCR. This revealed that N-gene specific PCR could achieve higher amplification efficiency than 1b-gene specific PCR when using same copy number of template. PCR with cDNA from 30 clinical samples or virus infected Vero cells were then performed. Figure 22A shows the Ct and half-maximal values of the fluorescent signal of N gene and 1b gene specific PCR.

generated from NPA, tracheal dispersion and lung biopsy from patient A. The results indicated that fluorescent signals given in N gene specific PCR are higher (26.0% in average, ranged from 6.3-60%) than that of 1b specific PCR in all positive samples.

Furthermore, Ct values of N gene specific PCR are lower (0.1-4.6 cycles) than that of 1b specific PCR among most of the SARS-CoV positive samples (Table 3).

Table 3

SN	X	Ib	? Ct	S/N	N	Ib	? Ct	S/N	N	Ib	? Ct	S/N	N	Ib	? Ct
55851	27.1	27.8	0.6	34852	31.9	33.2	1.3	67/29	35.3	35.6	0.3	68795	32.4	34.5	2.1
55751	27.0	27.7	0.1	45971	43.6	48.2	4.6	67/39	28.1	27.5	-0.9	69798	25.8	28.4	-0.4
62391	41.0	43.2	1.3	55972	42.6	46.5	2.9	65/18	30.0	33.5	3.5	68562	31.6	38.3	3.7
55331	41.0	43.1	2.1	55973	42.1	43.2	1.1	65/18	34.7	37.7	1.0	68801	31.9	32.8	0.9
55/X3	41.7	42.3	0.6	69145	27.2	27.7	0.5	68/34	52.4	33.2	0.8	70562	40.2	43.3	3.1
65733	43.6	44.9	1.3	56786	32.8	33.8	1.0	68/184	30.6	32.4	1.8	70589	35.5	38.2	2.7
34852	33.5	33.7	0.2	55527	37.3	39.4	2.1	68/185	27.5	30.1	2.6	70591	36.0	38.2	2.2
32N14	38.6	40.8	2.2	56851	23.9	26.1	2.2	68/187	40.3	41.5	1.2	70559	41.4	43.1	1.7
33935	35.3	36.5	1.2	69073	24.3	26.1	1.3	68/788	35.5	37.6	2.1				
34851	31.4	31.7	0.3	67423	28.7	29.4	0.7	68791	34.8	35.3	0.5				

$$\Delta Ct = 1.49 \pm 0.47, \text{ 95\% confidence intervals} = 0.74 \text{ to } 2.23 \text{ (F-test)}$$

10 Statistic analysis indicates that Ct of N-gene PCR assay is significantly lower than that of ab-gene assay (95% confidence interval = 0.74 to 2.23, F-test). Stronger fluorescence signals and lower Ct values of N gene specific PCR provide a more sensitive diagnostic result and much target for the assay.

Using cDNA from SARS-CoV infected Vero cells, amplification curves shown in
15 Fig. 21B show the differences between N gene and 1b gene specific PCR. Ct of the N gene and 1b gene specific PCR was 35.3 and 37.8, respectively. This phenomenon had two main causes: (1) Expression level of N gene was higher than that of 1b gene; and; (2) Copy number of N gene was much larger than that of 1b gene because each transcript preceded a copy of N gene, in SARS-CoV infected cells. Northern blot analysis supported this
20 hypothesis (Fig. 24). When N-gene specific PCR product was used as a probe, at least five transcripts from the virus were hybridized and gave positive signals (Figure 24). This result agreed with the findings in which five subgenomic mRNAs were detected by Northern hybridization of RNA from SARS-CoV infected cells using a probe derived from the 3' untranslated region (Rota *et al.*, 2003, *Science* 300:1394-99). On the other hand, when 1b
25 PCR product was used as a probe, only 2 transcripts with high molecular size were hybridized, demonstrating that the copy number of N gene was much higher than that of 1b

gene, during transcription and gene expression in the host cells. The Northern hybridization result strongly supports the conclusion that PCR amplifying regions in N gene of the SARS-CoV are more sensitive than other regions as a target for diagnostic screening. It is possible that amplification of more than one genome region may increase the specificity of the test

5 (Yam W.C., et al., 2003, *J. Clin. Microbiol.* 41:4521-24).

In conclusion, we have developed a new generation of RT-PCR diagnosis test which is more sensitive than conventional diagnostic test for the detection of the coronavirus associated with SARS. The assay provides a high throughput, highly sensitive screening platform, which enables us to scale up to test hundreds of thousands of suspected SARS 10 cases each day in a single working line. Incorporation of PK-15 cell as an internal control in the assay and use of N gene as a diagnosis locus in addition to 1b gene can enhance the sensitivity and accuracy of the test. We are adapting the protocol to 96-well real-time quantitative PCR and sequencing format to shorten the time required for the test and to obtain information on genotypic variation of the virus.

15

CLINICAL RESULTS

Clinical findings:

All 50 patients with SARS were ethnic Chinese. They represented 5 different epidemiologically linked clusters as well as additional sporadic cases fitting the case 20 definition. They were hospitalized at a mean of 5 days after the onset of symptoms. The median age was 42 years (range of 23 to 74) and the female to male ratio was 1.3. Fourteen (28%) were health care workers and five (10%) had a history of visit to a hospital experiencing a major outbreak of SARS. Thirteen (26%) patients had household contacts and 12 (24%) others had social contacts with patients with SARS. Four (8%) had a history 25 of recent travel to mainland China.

The major complaints from most patients were fever (90%) and shortness of breath. Cough and myalgia were present in more than half the patients (Table 4). Upper respiratory tract symptoms such as rhinorrhea (24%) and sore throat (20%) were present in a minority of patients. Diarrhea (10%) and anorexia (10%) were also reported. At initial examination, 30 auscultatory findings, such as crepitations and decreased air entry, were present in only 38% of patients. Dry cough was reported by 62% of patients. All patients had radiological evidence of consolidation, at the time of admission, involving 1 zone (in 36), 2 zones (13) and 3 zones (1).

Table 4

Clinical symptoms	Number (percentage)
Fever	50 (100%)
Chill or rigors	37 (74%)
Cough	31 (62%)
Myalgia	27 (54%)
Malaise	25 (50%)
Running nose	12 (24%)
Sore throat	10 (20%)
Shortness of breath	10 (20%)
Anorexia	10 (20%)
Diarrhea	5 (10%)
Headache	10 (20%)
Dizziness	6 (12%)

* Truncal maculopapular rash was noted in 1 patient.

5 In spite of the high fever, most patients (98%) had no evidence of a leukocytosis. Lymphopenia (68%), leucopenia (26%), thrombocytopenia (40%) and anemia (18%) were present in peripheral blood examination (Table 5). Parenchymal liver enzyme, alanine aminotransferase (ALT) and muscle enzyme, creatinine kinase (CPK) were elevated in 34% and 26% respectively.

Table 5

Laboratory parameter	Mean (range)	Percentage of abnormal	Normal range
Haemoglobin	12.9 (8.9 - 15.9)		11.5 - 16.5 g/dl
Anaemia		9 (18%)	
White cell count	5.17 (1.1 - 11.4)		4 - 11 x 10 ⁹ /L
Leucopenia		13 (26%)	
Lymphocyte count	0.78 (0.3 - 1.5)		1.5 - 4.0 x 10 ⁹ /L
Significant lymphopenia (<1.0 x 10 ⁹ /L)		34 (68%)	
Platelet count	174 (88 - 351)		150 - 400 x 10 ⁹ /L
Thrombocytopenia		20 (40%)	
Alanine aminotransaminase (ALT)	63 (11 - 350)		6 - 53 U/L
Elevated ALT		17 (34%)	
Albumin	37 (26 - 50)		42 - 54 g/L
Low albumin		34 (68%)	
Globulin	33 (21 - 42)		24 - 36 g/L
Elevated globulin		10 (20%)	
Creatinine kinase	244 (31 - 1379)		34 - 138 U/L
Elevated creatinine kinase		13 (26%)	

- Routine microbiological investigations for known viruses and bacteria by culture, antigen detection, and PCR were negative in most cases. Blood culture was positive for 5 *Escherichia coli* in a 74-year-old male patient, who was admitted to intensive care unit, and was attributed to hospital acquired urinary tract infection. *Klebsiella pneumoniae* and *Hemophilus influenzae* were isolated from the sputum specimens of 2 other patients on admission.
- Oral levofloxacin 500 mg q24h was given in 9 patients and intravenous (1.2 g q8h)/10 oral (375 mg tid) amoxicillin-clavulanate and intravenous/oral clarithromycin 500 mg q12h were given in another 40 patients. Four patients were given oral oseltamivir 75 mg bid. In one patient, intravenous ceftriaxone 2 gm q24h, oral azithromycin 500 mg q24h, and oral

amantadine 100 mg bid were given for empirical coverage of typical and atypical pneumonia.

Nineteen patients progressed to severe disease with oxygen desaturation and were required intensive care and ventilatory support. The mean number of days of deterioration 5 from the onset of symptoms was 8.3 days. Intravenous ribavirin 8 mg/kg q8h and steroid was given in 49 patients at a mean day of 6.7 after onset of symptoms.

The risk factors associated with severe complicated disease requiring intensive care and ventilatory support were older age, lymphopenia, impaired ALT, and delayed initiation of ribavirin and steroid (Table 6). All the complicated cases were treated with ribavirin and 10 steroid after admission to the intensive care unit whereas all the uncomplicated cases were started on ribavirin and steroid in the general ward. As expected, 31 uncomplicated cases recovered or improved whereas 8 complicated cases deteriorated with one death at the time of writing. All 50 patients were monitored for a mean of 12 days at the time of writing.

Table 6

	Complicated case (n= 19)	Uncomplicated case (n= 31)	P value
Mean (SD) age (range)	49.5 ± 12.7	39.0 ± 10.7	P < 0.01
Male / Female ratio	8 / 11	14 / 17	N.S.
Underlying illness	5 †	1 ‡	P < 0.05
Mode of contact			
Travel to China	1	3	N.S.
Health care worker	5	9	N.S.
Hospital visit	1	4	N.S.
Household contact	8	5	P < 0.05
Social contact	4	10	N.S.
Mean (SD) duration of symptoms to admission (days)	5.2 ± 2.0	4.7 ± 2.5	N.S.
Mean (SD) admission temperature (°C)	38.8 ± 0.9	38.7 ± 0.8	N.S.
Mean (SD) initial total peripheral WBC count ($\times 10^9 / L$)	5.1 ± 2.4	5.2 ± 1.8	N.S.
Mean (SD) initial lymphocyte count ($\times 10^9 / L$)	0.66 ± 0.3	0.85 ± 0.3	P < 0.05
Presence of thrombocytopenia (< $150 \times 10^9 / L$)	8	12	N.S.
Impaired liver function test	11	6	P < 0.01
CXR changes (number of zone affected)	1.4	1.2	N.S.
Mean (SD) day of deterioration from the onset of symptoms §	8.3 ± 2.6	Not applicable	
Mean (SD) day of initiation of Ribavirin & steroid from the onset of symptoms	7.7 ± 2.9	5.7 ± 2.6	P < 0.05
Initiation of ribavirin & steroid after deterioration	12	0	P < 0.001
Response to ribavirin & steroid	11	28	P < 0.05
Outcome			
Improved or recovered	10	31	P < 0.01
Not improving	8	0	P < 0.01

* Multi-varient analysis is not performed due to low number of cases;

† 2 patients had diabetic mellitus, 1 had hypertrophic obstructive cardiomyopathy, 1

5 had chronic active hepatitis B, and 1 had brain tumour;

‡ 1 patient had essential hypertension;

§ desaturation requiring intensive care support;

|| 1 died.

Two virus isolates, subsequently identified as a member of *Coronaviridae* (see below), were isolated from two patients. One was from an open lung biopsy tissue of a 53-year-old Hong Kong Chinese resident and the other from a nasopharyngeal aspirate of a 42-year-old female with good previous health. The 53-year old male had a history of 10-hour household contact with a Chinese visitor who came from Guangzhou and later died from SARS. Two days after this exposure, he presented with fever, malaise, myalgia, and headache. Crepitations were present over the right lower zone and there was a corresponding alveolar shadow on the chest radiograph. Hematological investigation revealed lymphopenia of $0.7 \times 10^9/l$ with normal total white cell and platelet counts. Both ALT (41 U/L) and CPK (405 U/L) were impaired. Despite a combination of oral azithromycin, amantadine, and intravenous ceftriaxone, there was increasing bilateral pulmonary infiltrates and progressive oxygen desaturation. Therefore, an open lung biopsy was performed 9 days after admission. Histopathological examination showed a mild interstitial inflammation with scattered alveolar pneumocytes showing cytomegaly, granular amphophilic cytoplasm and enlarged nuclei with prominent nucleoli. No cells showed inclusions typical of herpesvirus or adenovirus infection. The patient required ventilation and intensive care after the operative procedure. Empirical intravenous ribavirin and hydrocortisone were given. He succumbed 20 days after admission. In retrospect, coronavirus-like RNA was detected in his nasopharyngeal aspirate, lung biopsy and post-mortem lung. He had a significant rise in titer of antibodies against his own hSARS isolate from 1/200 to 1/1600.

The second patient from whom a hSARS virus was isolated, was a 42-year-old female with good past health. She had a history of travel to Guangzhou in mainland China for 2 days. She presented with fever and diarrhea 5 days after her return to Hong Kong. Physical examination showed crepitation over the right lower zone which had a corresponding alveolar shadow on the chest radiograph. Investigation revealed leucopenia ($2.7 \times 10^9/L$), lymphopenia ($0.6 \times 10^9/L$), and thrombocytopenia ($104 \times 10^9/L$). Despite the empirical antimicrobial coverage with amoxicillin-clavulanate, clarithromycin, and oseltamivir, she deteriorated 5 days after admission and required mechanical ventilation and intensive care for 5 days. She gradually improved without receiving treatment with ribavirin or steroid. Her nasopharyngeal aspirate was positive for the virus in the RT-PCR and she was seroconverted from antibody titre <1/50 to 1/1600 against the hSARS isolate.

Virological findings:

Viruses were isolated on FRhk-4 cells from the lung biopsy and nasopharyngeal aspirate respectively, of two patients described above. The initial cytopathic effect appeared between 2 and 4 days after inoculation, but on subsequent passage, cytopathic effect appeared in 24 hours. Both virus isolates did not react with the routine panel of reagents used to identify virus isolates including those for influenza A, B parainfluenza types 1,2,3, adenovirus and respiratory syncytial virus (DAKO, Glostrup, Denmark). They also failed to react in RT-PCR assays for influenza A and HMPV or in PCR assays for mycoplasma. The virus was ether sensitive, indicating that it was an enveloped virus.

Electron microscopy of negatively stained (2% potassium phospho-tungstate, pH 7.0) cell culture extracts obtained by ultracentrifugation showed the presence of pleomorphic enveloped viral particles, of about 80-90 nm (ranging 70-130 nm) in diameter, whose surface morphology appeared comparable to members of *Coronaviridae* (Figure 5a). Thin section electron microscopy of infected cells revealed virus particles of 55-90 nm diameter within the smooth-walled vesicles in the cytoplasm (Figure 5b). Virus particles were also seen at the cell surface. The overall findings were compatible with infections in the cells caused by viruses of *Coronaviridae*.

A thin section electron micrograph of the lung biopsy of the 53 year old male contained 60-90-nm viral particles in the cytoplasm of desquamated cells. These viral particles were similar in size and morphology to those observed in the cell-cultured virus isolate from both patients (Figure 4).

The RT-PCR products generated in a random primer RT-PCR assay were analyzed and unique bands found in the virus infected specimen was cloned and sequenced. Of 30 clones examined, a clone containing 646 base pairs (SEQ ID NO:1) of unknown origin was identified. Sequence analysis of this DNA fragment suggested this sequence had a weak homology to viruses of the family of *Coronaviridae* (data not shown). Deducted amino acid sequence (215 amino acids: SEQ ID NO:2) from this unknown sequence, however, had the highest homology (57%) to the RNA polymerase of bovine coronavirus and murine hepatitis virus, confirming that this virus belongs to the family of *Coronaviridae*. Phylogenetic analysis of the protein sequences showed that this virus, though most closely related to the group II coronaviruses, was a distinct virus (Figures 5a and 5b).

Based on the 646 bp sequence of the isolate, specific primers for detecting the new virus was designed for RT-PCR detection of this hSARS virus genome in clinical specimens. Of the 44 nasopharyngeal specimens available from the 50 SARS patients, 22

had evidence of hSARS RNA. Viral RNA was detectable in 10 of 18 fecal samples tested. The specificity of the RT-PCR reaction was confirmed by sequencing selected positive RT-PCR amplified products. None of 40 nasopharyngeal and fecal specimens from patients with unrelated diseases were reactive in the RT-PCR assay.

5 To determine the dynamic range of real-time quantitative PCR, serial dilutions of plasmid DNA containing the target sequence were made and subjected to the real-time quantitative PCR assay. As shown in Figure 7A, the assay was able to detect as little as 10 copies of the target sequence. By contrast, no signal was observed in the water control (Figure 7A). Positive signals were observed in 23 out of 29 serologically confirmed SARS 10 patients. In all of these positive cases, a unique PCR product ($T_m = 82^\circ\text{C}$) corresponds to the signal from the positive control was observed (Figure 7B, and data not shown). These results indicated this assay is highly specific to the target. The copy numbers of the target sequence in these reactions range from 4539 to less than 10. Thus, as high as 6.48×10^5 copies of this viral sequence could be found in 1 ml of NPA sample. In 5 of the above 15 positive cases, it was possible to collect NPA samples before seroconversion. Viral RNA was detected in 3 of these samples, indicating that this assay can detect the virus even at the early onset of infection.

To further validate the specificity of this assay, NPA samples from healthy individuals (n=11) and patients suffered from adenovirus (n=11), respiratory syncytial virus 20 (n=11), human metapneumovirus (n=11), influenza A virus (n=13) or influenza B virus (n=1) infection were recruited as negative controls. All of these samples, except one, were negative in the assay. The false positive case was negative in a subsequent test. Taken together, including the initial false positive case, the real-time quantitative PCR assay has sensitivity of 79% and specificity of 98 %.

25 Epidemiological data suggest that droplet transmission is one of the major route of transmission of this virus. The detection of live virus and the detection of high copies of viral sequence from NPA samples in the current study clearly support that cough and sneeze droplets from SARS patients might be the major source of this infectious agent. Interestingly, 2 out of 4 available stool samples from the SARA patients in this study were 30 positive in the assay (data not shown). The detection of the virus in feces suggests that there might be other routes of transmission. It is relevant to note that a number of animal coronaviruses are spread via the fecal-oral route (McIntosh K., 1974, Coronaviruses: a

comparative review. *Current Top Microbiol Immunol.* 63: 85-112). However, further studies are required to test whether the virus in feces is infectious or not.

Currently, apart from this hSARS virus, there are two known serogroups of human coronaviruses (229E and OC43) (Hruskova J. *et al.*, 1990, Antibodies to human 5 coronaviruses 229E and OC43 in the population of C.R., *Acta Virol.* 34:346-52). The primer set used in the present assay does not have homology to the strain 229E. Due to the lack of available corresponding OC43 sequence in the Genbank, it is not known whether these primers would cross-react with this strain. However, sequence analyses of available sequences in other regions of OC43 polymerase gene indicate that the novel human virus 10 associated with SARS is genetically distinct from OC43. Furthermore, the primers used in this study do not have homology to any of sequences from known coronaviruses. Thus, it is very unlikely that these primers would cross-react with the strain OC43.

Apart from the novel pathogen, metapneumovirus was reported to be identified in some of SARS patients (Center for Disease Control and Prevention, 2003, *Morbidity and 15 Mortality Weekly Report* 52: 269-272). No evidence of metapneumovirus infection was detected in any of the patients in this study (data not shown), suggesting that the novel hSARS virus of the invention is the key player in the pathogenesis of SARS.

Immunofluorescent antibody detection:

Thirty-five of the 50 most recent serum samples from patients with SARS had 20 evidence of antibodies to the hSARS (*see Fig. 3*). Of 27 patients from whom paired acute and convalescent sera were available, all were seroconverted or had >4 fold increase in antibody titer to the virus. Five other pairs of sera from additional SARS patients from clusters outside this study group were also tested to provide a wider sampling of SARS 25 patients in the community and all of them were seroconverted. None of 80 sera from patients with respiratory or other diseases as well as none of 200 normal blood donors had detectable antibody.

When either seropositivity to HP-CV in a single serum or viral RNA detection in the NPA or stool are considered evidence of infection with the hSARS, 45 of the 50 patients 30 had evidence of infection. Of the 5 patients without any virological evidence of *Coronaviridae* viral infection, only one of these patients had their sera tested > 14 days after onset of clinical disease.

DISCUSSION

The outbreak of SARS is unusual in a number of aspects, in particular, in the appearance of clusters of patients with pneumonia in health care workers and family contacts. In this series of patients with SARS, investigations for conventional pathogens of atypical pneumonia proved negative. However, a virus that belongs to the family *Coronaviridae* was isolated from the lung biopsy and nasopharyngeal aspirate obtained from two SARS patients, respectively. Phylogenetically, the virus was not closely related to any known human or animal coronavirus or torovirus. The present analysis is based on a 646 bp fragment (SEQ ID NO:1) of the polymerase gene, which indicates that the virus relates to antigenic group 2 of the coronaviruses along with murine hepatitis virus and bovine coronavirus. However, viruses of the *Coronaviridae* can undergo heterologous recombination within the virus family and genetic analysis of other parts of the genome needs to be carried out before the nature of this new virus is more conclusively defined (Holmes KV. Coronaviruses. Eds Knipe DM, Howley PM Fields Virology, 4th Edition, Lippincott Williams & Wilkins, Philadelphia, 1187-1203). The biological, genetic and clinical data, taken together, indicate that the new virus is not one of the two known human coronaviruses.

The majority (90%) of patients with clinically defined SARS had either serological or RT-PCR evidence of infection by this virus. In contrast, neither antibody nor viral RNA was detectable in healthy controls. All 27 patients from whom acute and convalescent sera were available demonstrated rising antibody titers to hSARS virus, strengthening the contention that a recent infection with this virus is a necessary factor in the evolution of SARS. In addition, all five pairs of acute and convalescent sera tested from patients from other hospitals in Hong Kong also showed seroconversion to the virus. The five patients who has not shown serological or virological evidence of hSARS virus infection, need to have later convalescent sera tested to define if they are also seroconverted. However, the concordance of the hSARS virus with the clinical definition of SARS appears remarkable, given that clinical case definitions are never perfect.

No evidence of HMPV infection, either by RT-PCR or rising antibody titer against HMPV, was detected in any of these patients. No other pathogen was consistently detected in our group of patients with SARS. It is therefore highly likely that that this hSARS virus is either the cause of SARS or a necessary pre-requisite for disease progression. Whether or

not other microbial or other co-factors play a role in progression of the disease remains to be investigated.

The family *Coronaviridae* includes the genus *Coronavirus* and *Torovirus*. They are enveloped RNA viruses which cause disease in humans and animals. The previously

- 5 known human coronaviruses, types 229E and OC43 are the major causes of the common cold (Holmes KV. Coronaviruses. Eds Knipe DM, Howley PM. *Fields Virology*, 4th Edition, Lippincott Williams & Wilkins, Philadelphia, pp. 1187-1203). But, while they can occasionally cause pneumonia in older adults, neonates or immunocompromised patient (El-Sahly HM, Atmar RL, Glezen WP, Greenberg SB. Spectrum of clinical illness in
10 hospitalized patients with "common cold" virus infections. *Clin Infect Dis*. 2000; **31**: 96-100; and Foltz EJ, Elkordy MA. Coronavirus pneumonia following autologous bone marrow transplantation for breast cancer. *Chest* 1999; **115**: 901-905), Coronaviruses have been reported to be an important cause of pneumonia in military recruits, accounting for up to 30% of cases in some studies (Wenzel RP, Hendley JO, Davies JA, Gwaltney JM,
15 Coronavirus infections in military recruits: Three-year study with coronavirus strains OC43 and 229E. *Am Rev Respir Dis*. 1974; **109**: 621-624). Human coronaviruses can infect neurons and viral RNA has been detected in the brain of patients with multiple sclerosis (Talbot PJ, Cote G, Arbour N. Human coronavirus OC43 and 229E persistence in neural cell cultures and human brains. *Adv Exp Med Biol*. - in press). On the other hand, a number
20 of animal coronaviruses (eg. Porcine Transmissible Gastroenteritis Virus, Murine Hepatitis Virus, Avian Infectious Bronchitis Virus) cause respiratory, gastrointestinal, neurological or hepatic disease in their respective hosts (McIntosh K. Coronaviruses: a comparative review. *Current Top Microbiol Immunol*. 1974; **63**: 85-112).

We describe for the first time the clinical presentation and complications of SARS.

- 25 Less than 25% of patients with coronaviral pneumonia had upper respiratory tract symptoms. As expected in atypical pneumonia, both respiratory symptoms and positive auscultatory findings were very disproportional to the chest radiographic findings. Gastrointestinal symptoms were present in 10%. It is relevant that the virus RNA is detected in faeces of some patients and that coronaviruses have been associated with diarrhoea in
30 animals and humans (Caul EO, Egglestone SI. Further studies on human enteric coronaviruses *Arch Virol*. 1977; **54**: 107-17). The high incidence of deranged liver function test, leucopenia, significant lymphopenia, thrombocytopenia and subsequent evolution into adult respiratory distress syndrome suggests a severe systemic inflammatory damage

induced by this hSARS virus. Thus immuno-modulation by steroid may be important to complement the antiviral therapy by ribavirin. In this regard, it is pertinent that severe human disease associated with the avian influenza subtype H5N1, another virus that recently crossed from animals to humans, has also been postulated to have an immuno-

- 5 pathological component (Cheung CY, Poon LLM, Lau ASY et al. Induction of proinflammatory cytokines in human macrophages by influenza A (H5N1) viruses: a mechanism for the unusual severity of human disease. *Lancet* 2002; 360: 1831-1837). In common with H5N1 disease, patients with severe SARS are adults, are significantly more lymphopenic and have parameters of organ dysfunction beyond the respiratory tract (Table
10 4) (Yuen KY, Chan PKS, Peiris JSM, et al. Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. *Lancet* 1998; 351: 467-471). It is important to note that a window of opportunity of around 8 days exists from the onset of symptoms to respiratory failure. Severe complicated cases are strongly associated with both underlying disease and delayed use of ribavirin and steroid therapy. Following our
15 clinical experience in the initial cases, this combination therapy was started very early in subsequent cases which were largely uncomplicated cases at the time of admission. The overall mortality at the time of writing is only 2% with this treatment regimen. There were still 8 out of 19 complicated cases who had not shown significant response. It is not possible to a detail analysis of the therapeutic response to this combination regimen due to
20 the heterogeneous dosing and time of initiation of therapy.

Other factors associated with severe disease is acquisition of the disease through household contact which may be attributed to a higher dose or duration of viral exposure and the presence of underlying diseases.

- The clinical description reported here pertains largely to the more severe cases
25 admitted to hospital. We presently have no data on the full clinical spectrum of the emerging *Coronaviridae* infection in the community or in an out-patient-setting. The availability of diagnostic tests as described here will help address these questions. In addition, it will allow questions pertaining to the period of virus shedding (and communicability) during convalescence, the presence of virus in other body fluids and excreta and the presence of virus shedding during the incubation period, to be addressed.

The epidemiological data at present appears to indicate that the virus is spread by droplets or by direct and indirect contact although airborne spread cannot be ruled out in some instances. The finding of infectious virus in the respiratory tract supports this

contention. Preliminary evidence also suggests that the virus may be shed in the feces. However, it is important to note that detection of viral RNA does not prove that the virus is viable or transmissible. If viable virus is detectable in the feces, this would be a potentially additional route of transmission that needs to be considered. It is relevant to note that a 5 number of animal coronaviruses are spread via the fecal-oral route (McIntosh K. Coronaviruses: a comparative review. *Current Top Microbiol Immunol.* 1974; 63: 85-112).

7. DEPOSIT

A sample of isolated hSARS virus was deposited with China Center for Type 10 Culture Collection (CCTCC) at Wuhan University, Wuhan 430072 in China on April 2, 2003 in accordance with the Budapest Treaty on the Deposit of Microorganisms, and accorded accession No. CCTCC-V200303, which is incorporated herein by reference in its entirety.

15 8. MARKET POTENTIAL

The hSARS virus can now be grown on a large scale, which allows the development of various diagnostic tests as described hereinabove as well as the development of vaccines and antiviral agents that are effective in preventing, ameliorating or treating SARS. Given the severity of the disease and its rapid global spread, it is highly likely that significant 20 demands for diagnostic tests, therapies and vaccines to battle against the disease, will arise on a global scale. In addition, this virus contains genetic information which is extremely important and valuable for clinical and scientific research applications.

9. EQUIVALENTS

25 Those skilled in the art will recognize, or be able to ascertain many equivalents to the specific embodiments of the invention described herein using no more than routine experimentation. Such equivalents are intended to be encompassed by the following claims.

All publications, patents and patent applications mentioned in this specification are herein incorporated by reference into the specification to the same extent as if each 30 individual publication, patent or patent application was specifically and individually indicated to be incorporated herein by reference.

Citation or discussion of a reference herein shall not be construed as an admission that such is prior art to the present invention.

WHAT IS CLAIMED.

1. An isolated nucleic acid molecule consisting essentially of a nucleic acid sequence of SEQ ID NO:2471, or a complement thereof.
2. An isolated nucleic acid molecule consisting essentially of a nucleic acid sequence of SEQ ID NO:2473, or a complement thereof.
3. An isolated nucleic acid molecule which hybridizes under stringent conditions to a nucleic acid molecule having the nucleotide sequence of claim 1 or 2, or a complement thereof.
4. The nucleic acid molecule of claim 1, 2 or 3 wherein the molecule is RNA.
5. The nucleic acid molecule of claim 1, 2 or 3 wherein the molecule is DNA.
6. An isolated nucleic acid molecule which hybridizes under stringent conditions to the nucleic acid molecule of claim 1 or 2, or a complement thereof, wherein the nucleic acid molecule encodes an amino acid sequence which has a biological activity exhibited by a polypeptide encoded by the nucleotide sequence of SEQ ID NO:2471 or 2473.
7. An isolated polypeptide encoded by the nucleic acid molecule of claim 1 or 2.
8. An antibody or an antigen-binding fragment thereof which immunospecifically binds to the N-gene protein of a hSARS virus.
9. An antibody or an antigen-binding fragment thereof which immunospecifically binds to the S-gene protein of a hSARS virus.
10. The antibody of claim 8, 9, or an antigen-binding fragment thereof which neutralizes the hSARS virus.
11. An antibody which immunospecifically binds to the polypeptide of claim 7, or an antigen-binding fragment of said antibody.
12. A method for detecting the presence of a N-gene of the hSARS virus in a biological sample, said method comprising:

- (a) contacting the sample with a compound that selectively binds to said N-gene; and
- (b) detecting whether the compound binds to said N-gene in the sample.

13. The method of claim 12, wherein the compound that binds to said N-gene is a nucleic acid molecule comprising a nucleotide sequence having at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150 or 1,200 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2471, or a complement thereof.

14. The method of claim 12, wherein the compound that binds to said N-gene is a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NO:2475, 2476, 2480 and/or 2481.

15. A method for detecting the presence of a S-gene of the hSARS virus in a biological sample, said method comprising:

- (a) contacting the sample with a compound that selectively binds to said S-gene; and
- (b) detecting whether the compound binds to said S-gene in the sample.

16. The method of claim 15, wherein the compound that binds to said S-gene is a nucleic acid molecule comprising a nucleotide sequence having at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 1,050, 1,100, 1,150, 1,200, 2,000, or 3,000 contiguous nucleotides of the nucleotide sequence of SEQ ID NO:2473, or a complement thereof.

17. The method of claim 15, wherein the compound that binds to said S-gene is a nucleic acid molecule comprising a nucleotide sequence of SEQ ID NO:2477 and/or 2478.

18. A method for detecting the presence of the polypeptide of claim 7 in a sample, said method comprising:

- (a) contacting the sample with a compound that selectively binds to said polypeptide; and

- (b) detecting whether the compound binds to said polypeptide in the sample.
19. The method of claim 18, wherein the compound that binds to the polypeptide is an antibody.
20. A method for identifying a subject infected with the hSARS virus, said method comprising:
- obtaining total RNA from a biological sample obtained from the subject
 - reverse transcribing the total RNA to obtain cDNA; and
 - subjecting the cDNA to real-time PCR assay using a set of primers derived from a nucleotide sequence of the N-gene of the hSARS.
21. The method of claim 20, wherein the set of primers have nucleotide sequences of SEQ ID NOS:2475 and 2476, respectively.
22. The method of claim 20, wherein the set of primers have nucleotide sequences of SEQ ID NOS:2480 and 2481, respectively.
23. A method for identifying a subject infected with the hSARS virus, said method comprising:
- obtaining total RNA from a biological sample obtained from the subject
 - reverse transcribing the total RNA to obtain cDNA; and
 - subjecting the cDNA to real-time PCR assay using a set of primers derived from a nucleotide sequence of the S-gene of the hSARS.
24. The method of claim 23, wherein the set of primers have nucleotide sequences of SEQ ID NOS:2477 and 2478, respectively.
25. A kit comprising in one or more containers one or more isolated nucleic acid molecules comprising a nucleotide sequence of SEQ ID NO:2475 and/or SEQ ID NO:2476.
26. A kit comprising in one or more containers one or more isolated nucleic acid molecules comprising a nucleotide sequence of SEQ ID NO:2480 and/or SEQ ID NO:2481.

27. A kit comprising in one or more containers one or more isolated nucleic acid molecules comprising a nucleotide sequence of SEQ ID NO:2477 and/or SEQ ID NO:2478.
28. A kit comprising in one or more containers one or more antibodies of claim 8 or 9.
29. A kit comprising in one or more containers one or more antibodies of claim 11.

1/106

a cag gac gct gta gct tca aaa atc tta gga ttg cct acg cag act gtt 49
 Gln Asp Ala Val Ala Ser Lys Ile Leu Gly Leu Pro Thr Gln Thr Val
 1 5 10 15
 gat tca tca cag ggt tct gaa tat gac tat gtc ata ttc aca caa act 97
 Asp Ser Ser Gln Gly Ser Glu Tyr Asp Tyr Val Ile Phe Thr Gln Thr
 20 25 30
 act gaa aca gca cac tct tgt aat gtc aac cgc ttc aat gtg gct atc 145
 Thr Glu Thr Ala His Ser Cys Asn Val Asn Arg Phe Asn Val Ala Ile
 35 40 45
 aca agg gca aaa att ggc att ttg tgc ata atg tct gat aga gat ctt 193
 Thr Arg Ala Lys Ile Gly Ile Leu Cys Ile Met Ser Asp Arg Asp Leu
 50 55 60
 tat gac aaa ctg caa ttt aca agt cta gaa ata cca cgt cgc aat gtg 241
 Tyr Asp Lys Leu Phe Thr Ser Leu Glu Ile Pro Arg Arg Asn Val
 65 70 75 80
 gct aca tta caa gca gaa aat gta act gga ctt ttt aag gac tgt agt 289
 Ala Thr Leu Gln Ala Glu Asn Val Thr Gly Leu Phe Lys Asp Cys Ser
 95 90 95
 aag atc att act ggt ctt cat cct aca gag gca cct aca cac ctc agc 337
 Lys Ile Ile Thr Gly Leu His Pro Thr Gln Ala Pro Thr His Leu Ser
 100 105 110
 gtt gat ata aaa ttc aag act gaa gga tta tgt gtt gac ata cca ggc 385
 Val Asp Ile Lys Phe Lys Thr Glu Gly Leu Cys Val Asp Ile Pro Gly
 115 120 125
 ata cca aag gac atg acc tac cgt aca ctc atc tct atg atg ggt ttc 433
 Ile Pro Lys Asp Met Thr Tyr Arg Arg Leu Ile Ser Met Met Gly Phe
 130 135 140
 aaa atg aat tac caa gtc aat ggt tac cct aat atg ttt atc acc cgc 481
 Lys Met Asn Tyr Gln Val Asn Gly Tyr Pro Asn Met Phe Ile Thr Arg
 145 150 155 160
 gaa gaa gct att cgt cac gtt cgt gcg tgg att ggc ttt gat gta gag 529
 Glu Glu Ala Ile Arg His Val Arg Ala Trp Ile Gly Phe Asp Val Glu
 165 170 175
 ggc tgt cat gca act aga gat gct gtg ggt act aac cta cct ctc cag 577
 Gly Cys His Ala Thr Arg Asp Ala Val Gly Tha Asn Leu Pro Leu Gln
 180 185 190
 cta gga ttt tct aca ggt gtt aac tta gta gct gta ccg act ggt tat 625
 Leu Gly Phe Ser Thr Gly Val Asn Leu Val Ala Val Pro Thr Gly Tyr
 195 200 205
 gtt gac act gaa aat aac cta 646
 Val Asp Thr Glu Asn Asn Leu
 210 215

FIG. 1

2/106

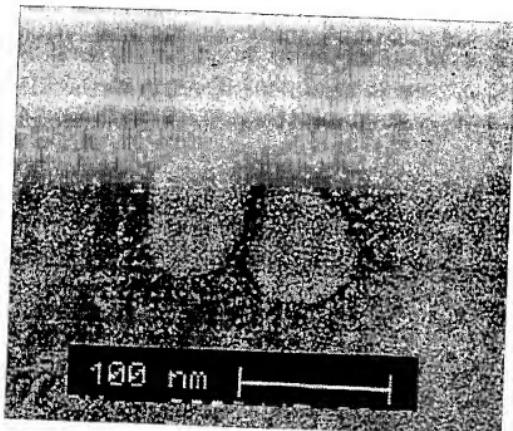


FIG. 2

3/106



FIG. 3

4/106

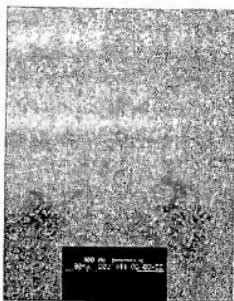


FIG. 4

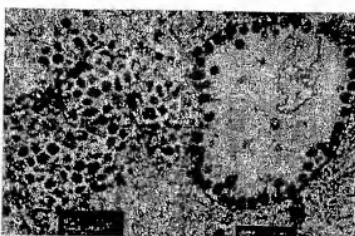


FIG. 5A

FIG. 5B

5/106

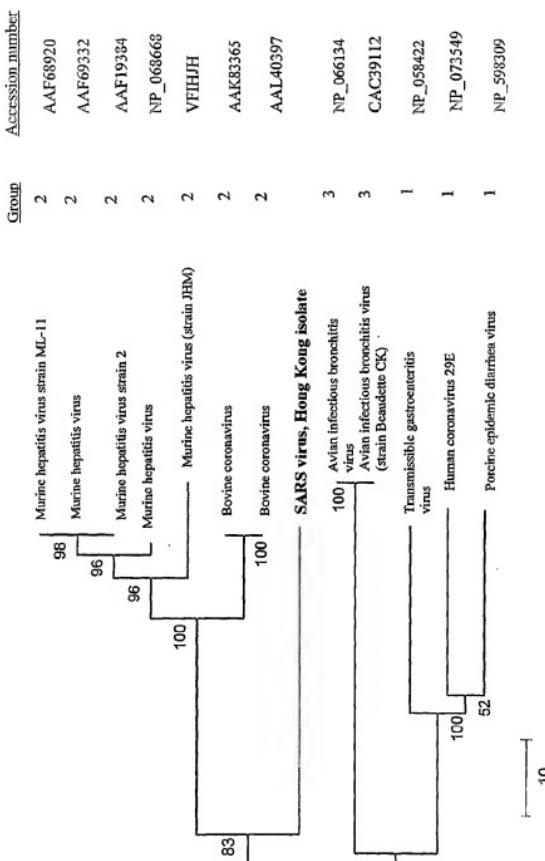


FIG. 6

6 / 106

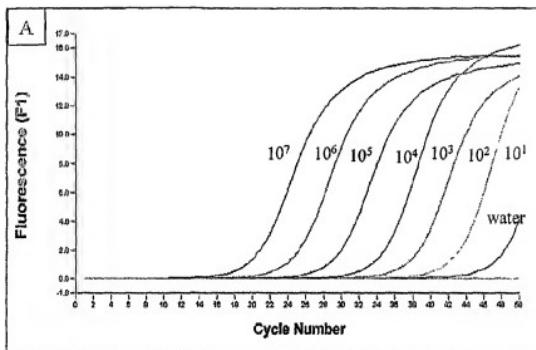


FIG. 7A

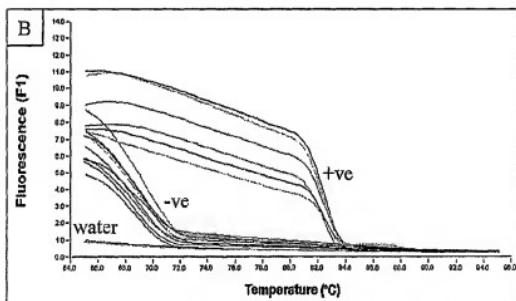


FIG. 7B

7/106

t aaa tgt agt aga atc ata cct gog cgt gog cgc gta gag tgt ttt gat	49
Lys Cys Ser Arg Ile Ile Pro Ala Arg Ala Arg Val Glu Cys Phe Asp	
1 5 10 15	
aaa ttc aac gtg aat tca aca cta gaa cag tat gtt ttc tgc act gta	97
Lys Phe Lys Val Asn Ser Thr Leu Glu Gln Tyr Val Phe Cys Thr Val	
20 25 30	
aat gca ttg cca gaa aca act gct gac att gta gtc ttt gat gaa atc	145
Asn Ala Leu Pro Glu Thr Thr Ala Asp Ile Val Val Phe Asp Glu Ile	
35 40 45	
tct atg gct act aat tat gac ttg agt gtt gtc aat gct aca ctt cgt	193
Ser Met Ala Thr Asn Tyr Asp Leu Ser Val Val Asn Ala Arg Leu Arg	
50 55 60	
gca aca ccc tac gtc tat att ggc gat cct gct cca tta cca gcc ccc	241
Ala Lys His Tyr Val Tyr Ile Gly Asp Pro Ala Gln Leu Pro Ala Pro	
65 70 75 80	
cgc aca ttg ctg act aac ggc aca cta gaa cca gaa tat ttt aat tca	289
Arg Thr Leu Leu Thr Lys Gly Thr Leu Glu Pro Glu Tyr Phe Asn Ser	
85 90 95	
gtg tgc aca ctt atg aac aca ata ggt cca gac atg ttc ctt gga act	337
Val Cys Arg Leu Met Lys Thr Ile Gly Pro Asp Met Phe Leu Gly Thr	
100 105 110	
tgt cgc cgt tct cct gct gaa att gtt gac act gtg agt gct tta gtt	385
Cys Arg Cys Pro Ala Glu Ile Val Asp Thr Val Ser Ala Leu Val	
115 120 125	
tat gac aat aag cta aac gca cac aag gag aag tca gct cca tgc ttc	433
Tyr Asp Asn Lys Leu Lys Ala His Lys Glu Lys Ser Ala Gln Cys Phe	
130 135 140	
aaa atg ttc tac aac ggt gtt att aca cat gat gtt tca tct gca atc	481
Lys Met Phe Tyr Lys Gly Val Ile Thr His Asp Val Ser Ser Ala Ile	
145 150 155 160	
aac aga cct caa ata ggc gtt gta aga gaa ttt ctt aca cgc aat cct	529
Asn Arg Pro Gln Ile Gly Val Val Arg Glu Phe Leu Thr Arg Asn Pro	
165 170 175	
gct tgg aga aac gct gtt ttt atc tca cct tat aat tca cag aac gct	577
Ala Trp Arg Lys Ala Val Phe Ile Ser Pro Tyr Asn Ser Gln Asn Ala	
180 185 190	
gta gct tca aac atc tta gga ttg cct acg cag act gtt gat tca tca	625
Val Ala Ser Lys Ile Leu Gly Leu Pro Thr Gln Thr Val Asp Ser Ser	
195 200 205	
aac ggt tct gaa tat gac tat gtc ata ttc aca caa act act gaa aca	673
Gln Gly Ser Glu Tyr Asp Tyr Val Ile Phe Thr Gln Thr Glu Thr	
210 215 220	

FIG. 8

8/106

gca ccc tct tgt aat gtc aac cgc ttc aat gtg gct atc aca agg gca 721
 Ala His Ser Cys Asn Val Asn Arg Phe Asn Val Ala Ile Thr Arg Ala
 225 230 235 240
 aaa att ggc att ttg tgc ata atg tct gat aga gat ctt tat gag aas 769
 Lys Ile Gly Ile Leu Cys Ile Met Ser Asp Arg Asp Leu Tyr Asp Lys
 245 250 255
 ctg cca ttt aca agt cta gaa ata cca cgt cgc aat gtg gct aca tta 817
 Leu Gln Phe Thr Ser Leu Glu Ile Pro Arg Arg Asn Val Ala Thr Leu
 260 265 270
 caa gca gaa aat gta act gga ctt ttt aag gac tgt agt aag atc att 865
 Gln Ala Glu Asn Val Thr Gly Leu Phe Lys Asp Cys Ser Lys Ile Ile
 275 280 285
 act ggt ctt cat cct aca cag gca cct aca cac ctc agc gtt gat ata 913
 Thr Gly Leu His Pro Thr Gln Ala Pro Thr His Leu Ser Val Asp Ile
 290 295 300
 aaa ttc aag act gaa gga tta tgt gtt gac ata cca ggc ata cca aag 961
 Lys Phe Lys Thr Glu Gly Leu Cys Val Asp Ile Pro Gly Ile Pro Lys
 305 310 315 320
 gac atg acc tac cgt aga ctc atc tot atg atg ggt ttc aaa atg aat 1009
 Asp Met Thr Tyr Arg Leu Ile Ser Met Met Gly Phe Lys Met Asn
 325 330 335
 tac caa gtc sat ggt tac cct aat atg ttt atc acc cgc gaa gct 1057
 Tyr Gln Val Asn Gly Tyr Pro Asn Met Phe Ile Thr Arg Glu Glu Ala
 340 345 350
 att cgt ccc gtt cgt ccc tgg att ggc ttt gat gta gag ggc tgt cat 1105
 Ile Arg His Val Arg Ala Trp Ile Gly Phe Asp Val Glu Gly Cys His
 355 360 365
 gca act aga gat gct gtg ggt act aac cta cct ctc cag cta gga ttt 1153
 Ala Thr Arg Asp Ala Val Gly Thr Asn Leu Pro Leu Gln Leu Gly Phe
 370 375 380
 tct aca ggt gtt aac tta gta gct gta cog act ggt tat gtt gac act 1201
 Ser Thr Gly Val Asn Leu Val Ala Val Pro Thr Gly Tyr Val Asp Thr
 385 390 395 400
 gaa aat aac cta 1213
 Glu Asn Asn Leu

FIG. 8 Con't

9/106

c aga acc atg cct aac atg ctt agg ata atg gcc tct ctt gtt ctt gct 49
 Arg Thr Met Pro Asn Met Leu Arg Ile Met Ala Ser Leu Val Leu Ala
 1 5 10 15

cgc aaa cat aac act tgc tgt aac tta tca cac cgt ttc tac agg tta 97
 Arg Lys His Asn Thr Cys Cys Asn Leu Ser His Arg Phe Tyr Arg Leu
 20 25 30

gct aac gag tgt gcg caa gta tta aat gag atg gtc atg tgt ggc ggc 145
 Ala Asn Glu Cys Ala Gln Val Leu Ser Glu Met Val Met Cys Gly Gly
 35 40 45

tca ctc tat gtt aat cca ggt gga aca tca tcc ggt gat gct aca act 193
 Ser Leu Tyr Val Lys Pro Gly Thr Ser Ser Gly Asp Ala Thr Thr
 50 55 60

gct tat got aat aat gtc ttt aac att tgt caa gca gtt aat gct aca act 241
 Ala Tyr Ala Asn Ser Val Phe Asn Ile Cys Gln Ala Val Thr Ala Asn
 65 70 75 80

gta aat gca ctt ctt tca act gat ggt aat aag ata gct gac aag tat 289
 Val Asn Ala Leu Ser Thr Asp Gly Asn Lys Ile Ala Asp Lys Tyr
 95 90 95

gtc cgc aat cta caa cac agg ctc tat gag tgt ctc tat aca aat aag 337
 Val Arg Asn Leu Gln His Arg Leu Tyr Glu Cys Leu Tyr Arg Asn Arg
 100 105 110

gat gtt gat cat gaa ttc gtg gat gag ttt tac gct tac ctg cgt aat 385
 Asp Val Asp His Glu Phe Val Asp Glu Phe Tyr Ala Tyr Leu Arg Lys
 115 120 125

cat ttc tcc atg atg att ctt tct gat gat gcc gtt gtg tgc tat aac 433
 His Phe Ser Met Met Ile Leu Ser Asp Asp Ala Val Val Cys Tyr Asn
 130 135 140

agt aac tat gcg gct caa ggt tta gta gct aca att aag aac ttt aag 481
 Ser Asn Tyr Ala Ala Gln Gly Leu Val Ala Ser Ile Lys Asn Phe Lys
 145 150 155 160

gca gtt ctt tat tat csa aat aat gtg ttc atg tct gag gca aaa tgt 529
 Ala Val Leu Tyr Glu Asn Asn Val Phe Met Ser Glu Ala Lys Cys
 165 170 S 175

tgg act gag act gac ctt act aat gga ccc cac gaa ttt tgc tca cag 577
 Trp Thr Glu Thr Asp Leu Thr Lys Gly Pro His Glu Phe Cys Ser Gln
 180 185 190

cat aca atg cta gtt aat csa gga gat gat tac gtg tac ctg cct tac 625
 His Thr Met Leu Val Lys Gln Gly Asp Asp Tyr Val Tyr Leu Pro Tyr
 195 200 205

cca gat cca tca aga aat tta ggc gca ggt tgt ttt gtc gat gat att 673
 Pro Asp Pro Ser Arg Ile Leu Gly Ala Gly Cys Phe Val Asp Asp Ile
 210 215 220

gtc aaa cag atg gta cac tta tga ttg aat ggt tcc gtg tca ctg gct 721
 Val Lys Gin Met Val His Leu
 225 230

att gat gc 729

FIG. 9

10/106

FIG. 10

11/106

FIG. 10 Con't

12/106

FIG. 10 Con't

13/106

10081 gatgacacag tatactgtcc aagacatgtc atttgcacag cagaagacat gcttaatct
 10141 aacctatgaq atctgcgtcal tgccaaatcc accatacatgt ttcttgcttc ggctggccat
 10201 gttcaactt gtgttatttg ccattctatg caaaaatgtc tgcttaggt aaagtgtat
 10261 aetcctaaacc ctaagacacc caagatataaa ttgtcccgta tccaaacctgg tcaaaccatt
 10321 tcgtgtctcg catgtccaa tggtccacca ttgtgtgtt atcaatgtcg catgagacat
 10381 aatcatatcca ttaaagggtt ttcccttatg gatcatgtg tgatgttgq tttaaacatt
 10441 gatgtatgatt ggtgtgttt ctgtatatacgttcatatgg agcttccaa acggatcac
 10501 gctggacty actttagaagg taaatctatg gtcatttgc ttgacgacat aactgcac
 10561 gctgcggatgta cagacacacaa cataacatta aatgttttg catggctgtb tgctgtctt
 10621 atcaatgtgtt atagggtgtt ttcttaatga ttcaacacta ctgttgatgt ctttaacatt
 10681 gtggcaatgt aqatacaacta tgaaaccttg aacaagatc atgtgtacat atggggact
 10741 cttttgcctc aaacaaaggat tgccgttca gatatgtgtg ctgtttggaa agagctgt
 10801 cagaatgtta tgaatgtgtg tactatccc gtagacta tttttagaaga tgatgttaca
 10861 ccattttggc ttgttagatca atgttgtgtt ctgttccaa aaggttaatg caagaaatt
 10921 gtttaagggtca cttcatatgg tggcttttgc actttttgtt gatcttgcataat gatctgtt
 10981 caaagtcaac atgtgtcaact gttttttttt gtttacgaga atgttcttcc gccattttact
 11041 ctgggtata tggcaatgt tgcatgtgtg atgtgtgtt ttaaqtatca gcaacgttcc
 11101 ttgtgtctgt ttctgttacc ttcttcttgoa acatgtgttcc actttatat ggtctcatg
 11161 ctgtgtactg ggggtatgtc tatcatgacaa tggttgaat tggtgtacat tagttgtct
 11221 ggttataatggc tttaagggtt tggtatgtat gtttcaatgtt tagtttgc tatctctatg
 11281 acatgtcgcg ctgtttatgtgatgtgttcatgtgttgg gtagactgtt ggacactgtt gaatgttcat
 11341 acatctgttca aaaaatgtca ctatgttaat gttttagatc aatgttatttc ctatgtggcc
 11401 ttatgttattt ctgttacactc taatcttattt ggtgtgttgc cgatctatcat gttttatgt
 11461 aagatgtatg tgttgtgtg tggtttagatgtaatccatttttattttt tggceacacc
 11521 ttacatgtatg tcaatgtgtt ttatgtttttt tgtaggtttt gttgtgtgtt ctatcttgc
 11581 cttttcttgc tacccaaacc ttacttccgg ctatcttgc ttgtttatgaat ctatctgt
 11641 ttacacaaag aattttatgtgatgtacttcc caggggtttt tgccctctaa gagtagtatt
 11701 gatgtttttt acgtttatgtt gttttttttt gttttttttt tgtaaaaaatg tttttttttt
 11761 gctactgttca atgttcaaat gttttttttt gttttttttt gttttttttt gttttttttt
 11821 ctccaaacaaat tttagatgtt gttttttttt aatttttttt gttttttttt gttttttttt
 11881 aatgtatata tttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 11941 btgtgttgc tttttttttt ggtgtgttca gacatataat gttttttttt gttttttttt
 12001 gataaccgtt ctatcttca ggttatttttccatgtt gttttttttt gttttttttt
 12061 gtttcatgccc ctggcccgaa ggccatgttgc cttttttttt gttttttttt gttttttttt
 12121 gtttccaaatg atttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 12181 gctttttttt gaaatgttgg gaaatgttgg gaaatgttgg gaaatgttgg gaaatgttgg
 12241 gcaagatgtg aggacaaatgg gcaaaatggt acttgcgttca tggcaacatctt gttttttttt
 12301 atgtttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 12361 ttgtgttccaa cccatcatgtt accatgttactt acagacggca aactcatgtt gttttttttt
 12421 gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 12481 ttggaaatcc acgtttatgtt gttttttttt gttttttttt gttttttttt gttttttttt
 12541 atggacaaatcc cccatcatgtt accatgttactt acagacggca aactcatgtt gttttttttt
 12601 gctttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 12661 gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 12721 aagggtggatg gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 12781 ttcccttaatg ttgtgtgttccatgtt accatgttactt acagacggca aactcatgtt gttttttttt
 12841 gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 12901 aacccatataat gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 12961 aatgttactatg aatgttactatg gttttttttt gttttttttt gttttttttt gttttttttt
 13021 cctgtttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 13081 aatgtttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 13141 atggacaaatcc acgtttatgtt gttttttttt gttttttttt gttttttttt gttttttttt
 13201 catccaaatcc cttaaaggat gttttttttt gttttttttt gttttttttt gttttttttt
 13261 ttgtgttccatgtt accatgttactt acagacggca aactcatgtt gttttttttt gttttttttt
 13321 tgaaatgtt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt
 13381 gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt gttttttttt

FIG. 10 Com't

14 / 106

FIG. 10 Con't

15/106

16801 gaggtactac gacatacaag ttgaatgtt gtgttactt tggatgtaca totcacactg
 16861 taatgcact tagtgcacct actcttagtgc cacaaggaoa ctatgtgaga attactgtc
 16921 tgatccaaact actcaacatc tcatacgat ttcttagca tggatgtcaat tatcaaaagg
 16981 tggcatacgca aaagtactct acactccaa gaccacactg tactgtgtaa agtcatttt
 17041 ccatacgact tgcctctatc taccatctcg ctgcatacgat gtatacgcc tgctctatc
 17101 cagctgttgc tgcocatgtg gaaaaggcat taaaatattt gcccatacatg aatgtatgt
 17161 gaatoatacc tggcgtgtcg cgcgtatgtg gttttgatata attcaaaatg aattcaaac
 17221 tagaaacgta tggtttctgtc actgttaatgt cattggcaga aacaaactgtc gacatgttg
 17281 totttgtatc aatcttctatc gacttaattt atgactgtgg tggatgtcaat gotagacttc
 17341 gtgcacaaaca ctacgttcatl atttgcgatc ctgtcaattt accacgcccc cgcacatgtc
 17401 tgactcaaaagg cacatcttgc coaagatattt ttaatccatgt gtgcacatgtc atgaaacaa
 17461 taggtccaga ctatgttctt ggcactttt ggcgtgttgc tggtgaaattt gtgtacatgt
 17521 tggatgttgc ttgttatgtc aataagctaa aacgcacacaa ggataatgtca gtcatactgt
 17581 tcaaaatgtt gtttttttttgc atgtatgttgc atgtatgttgc atgtatgttgc aacagactt
 17641 aataeggtt gtttttttttgc gtttttttttgc gtttttttttgc ttggaaaaaaat gttttttt
 17701 totccatcca taatccgc aacgcgttgc ctccaaaaat tttaggatgttgc cotacgdcga
 17761 ctgttgcattc atcacagggtt ctgtatgtt actatgttgc attacacacaa actactgtaaa
 17821 cagcacacto ttgtatgttgc aacccgttca atgttgcattt cacaaggccca aaaaatgyca
 17881 ttttgcattt aatgttgcattt agatgttgcattt atgacaaactt gcaatttaca agtotaaaaa
 17941 tacccacgtcg caatgttgcattt acattacaaag cagaaaaatgt aactggactt ttaaaggact
 18001 tggatgtatc acttacttgc ttccatctca caccaggccac tacacacccctc acgtgttgc
 18061 taaaatccatc gactgttgcattt ttatgttgc attacacaaag cataccaaaag gacatgttgc
 18121 accgttagact ctttgcattt atgggtttca aatgttgcattt ccaatgtcaat ggtttgcattt
 18181 atatgttttgc caccggcgaa gaaatgttgc ttccatctca atgttgcattt
 18241 tagggcgccgcttgc atgatgttgcattt agatgttgcattt ttggatgttgcattt cccatcttgc
 18301 ttccatctca gtttttttttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 18361 cagaatccatc caggttttttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 18421 cactctatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 18481 gtgtatcatc gaaaggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 18541 ggcttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 18601 acaaaatgttgc aacttgcattt ttccatctca ctttgcattt ttgtatgttgc ttgtatgttgc
 18661 tgggttttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 18721 gtaacccatc gatgttgcattt gcaacacatc gocatgttgc ttgtatgttgc ttgtatgttgc
 18781 ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 18841 attttttttttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 18901 aatgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 18961 acatgttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19021 acatgttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19081 ttccatctca ctttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19141 ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19201 taccaggcgccg ttttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19261 tccatctca gatgttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19321 ctgttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19381 ctatgttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19441 accgttagactt ctgttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19501 acaaaatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19561 atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19621 ttccatctca gatgttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19681 aaaataagatc aacacttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19741 aaccatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19801 taatotgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19861 tgactgtatc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19921 atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 19981 cagaaggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 20041 gactgttgcattt atggatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc
 20101 gtttttttttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc ttgtatgttgc

FIG. 10 Com't

16/106

20161 agccccgatc acaaatggaa actgacttto tgtagctcg tatggatgaa ttcatacagc
20221 gataatacggt ccggaggatc ggcttgcggcc acatggttt tggggatcc agtcgtac
20281 actatggcc tccttcatttc atgatggctc tagccaaagg ctccaaatgt ttccatcca
20341 aataggaga ttttccatcc atggacggcc cggatggaaa ttatccatc acatggatc
20401 aasaggctt ctttttttttcc aatggaaatgt tggtgttgc tgatgttgc ttactatc gatgttgcc
20461 agatataaa gtcaacaatgt ttgttgcgtga ttccaaatgg gtcggatgtt gtcggatgtt aatcatgg
20521 atgtatggat ttttttttttcc atggatggcc ttgttttttttcc aggtggatca tggttttttttcc
20581 aactacaacg aagtcaacgg tggcaaccgg ggttgtccat gcttaacttgc tcaaaatgtc
20641 aaagaatgtc ttgttggaaatgg tggtggatcc aggtttatgt tggaaatgtc gtatccaaatgt
20701 aagaaatgtc gagatgttgc gaaatggatc ctcaatgttgc tcaatcttca aatcatatca
20761 ctttggatcc accatccatc atggatggta ttccatccgt tggttttttttcc
20821 ttttttttttcc tacagctgtg ctccatggatc gttgttgcacca tgccatccatc ttgttgcatt
20881 catggatgtc tgatgttgc tccggatggcc attttacttgc aatggatggatc ttatgttgc
20941 taatccatgg taataatggt gacccatcc ttggatggatc tgatgttgc
21001 atgtgacaaaa agagaatgtac tcttccaaatgg ggttttttttcc
21061 aaccaaaaactt acggccgttgtt ggttttttttcc
21121 ctggatccca caatggatcc gggccatccatc ttggatggatc
21181 atgtatccatc atggatggatc tttttttttcc
21241 aatggatgtc ctatccatgg ttttttttttcc
21301 atgttgcgtt ctatccatgg ttttttttttcc
21361 ctgtatggatc ttttttttttcc
21421 tgtagcttatc ctttttttttcc
21481 acttacaaatgg attttttttcc
21541 acggccgttcc ctttttttttcc
21601 tgagggggggtt ttttttttttcc
21661 atttttttttcc ttttttttttcc
21721 gaaacccctgtt ctttttttttcc
21781 ttgttccgttgg ttttttttttcc
21841 ttaacatccatc ttttttttttcc
21901 ttgttccgttgg ttttttttttcc
21961 ttatggatccatc ttttttttttcc
22021 gtatatttttca acatccatgg ttttttttttcc
22081 ataaagggttcc ttttttttttcc
22141 aaccttttttcc ttttttttttcc
22201 ctttttttttcc ttttttttttcc
22261 taaggccaaatc ttttttttttcc
22321 atgttccatc aatccatggatcc ttttttttttcc
22381 aagaaatgttcc ttttttttttcc
22441 ctatataatgg ttttttttttcc
22501 ttatgtatcc gggaaaaatccatc ttttttttttcc
22561 acttacaaatccatc ttttttttttcc
22621 ttgttccgttgg ttttttttttcc
22681 acaactgttcc ttttttttttcc
22741 tggttccgttgg ttttttttttcc
22801 atataataatggatcc ttttttttttcc
22861 atgttccatc cttccatgg ttttttttttcc
22921 cttatataatgg ttttttttttcc
22981 tagtacttcc ttttttttttcc
23041 ctggatccatc ttttttttttcc
23101 ttgttccatcc ttttttttttcc
23161 atttccatcc ttttttttttcc
23221 ttgttccgttgg ttttttttttcc
23281 atgttccatcc ttttttttttcc
23341 cagcttggcc ttttttttttcc
23401 tagggatccatc ttttttttttcc
23461 ttgttccatcc ttttttttttcc

FIG. 10 Con't

17/106

23521 atatactatgtc tttagtgctg gatagttcaa ttgttactc taataacaco attgtctatac
 23581 ctactaactt ttcataattgc attactacag aagtaatgco tggttctatg gatcaaaact
 23641 cggtagatt taatatgtac attgtcgccg atctactgc atgtgtcaat ttgttctcc
 23701 aatatgtatg ctttgcacca caactaaato gtgcacttc aggtatgtt gctgcaacagg
 23761 atccgaacac acgtgaatg ttgcotcaag tcaacaaaat gtacaaacc ccaacttiga
 23821 aatattttgg tggttttawt tttcacaaaaa tatttactga ccctctaaag ccaactaaga
 23881 ggctttttal ttaggactctt ttgttataata aggtgacact cgctgtatgtt ggttcaatga
 23941 acaaatatgg cgaatgttca ggttatata atgtctagat ttctatgtt gogcaatgg
 24001 tcaatgtact tacagtgttg ccacacttgc tcactgtga tatgtatgtt goctacactg
 24061 ctgtctatgt tagtggact gcaactgtc gatggacat gatgtctgg gatgtcttcc
 24121 aaataccatcc ttgttcatgtt atggcaatgg gtttcaatgg catgtggatg accaaaaatg
 24181 ttcttctatgtt gaaacaaaaa caaatccca accaaattna caaggcattt agtcaatcc
 24241 aagaatcaact tacacaacaa tcaactgtat tggcaatgg gcaagacgtt gtaaccagg
 24301 atgttcaago attaaadcaat cttgttaaac aacttagctc taatttttgtt gaaatttca
 24361 ttgttctaaat tgatattctt tggcacttgc ataaatgtca ggccggatgtt caaatggca
 24421 gtttataatgg aggccatgtt caaaggcttta aacactatgtt aacacaacaa caatccagg
 24481 ctgtgttaat cagggttcc tgotaatottt ctgtatctat aatgtctgtg tggttcttgg
 24541 gacatcaaaa aagttgttgc tttttgtggaa gttgttccacc ctttatgtcc ttccccaaac
 24601 cagcccccga tggtgttgc ttccatcatg tcaactgtt gatccatcccg gagagaaact
 24661 tcaaccacage gccaatcaat tgcataatgg gcaaaatggat cttccctcgtl gaaagggttt
 24721 ttgttgttgc ttggcacttgc tggtttataa caacaggata ttctttttt ccaccaataa
 24781 ttactacaga caatacattt gttctcaaggaa atgtgtatgtt ctgtatttgc atcatataa
 24841 acacatgtt tgatcttgc caaccttgc tggactatccat caaaggatgg cttggcaatgg
 24901 acttccaaaaa ctatccatca ccaatgttgc atcttgcgtt catttgcggc attaaaccc
 24961 ctgtcgctca cattttttttt gaaatgttggc gtttcaatgtg gtttgcctaa aattttaaatg
 25021 aatactcaat tgcatttcaaa gaaatggggaa atatgtatca atatattaaa tggcttttt
 25081 atgttttgttggt oggttcttgc tggttggactt ttgcocatgtt catgttaca atcttgtt
 25141 gtttgcatgtc tagttttgc agtttgccttca ggggttgcat ctcttttgtt ttttgttgc
 25201 atgttgcattt ggtatgttgc gggccatgtt tcaagggtgtt caaatatcat tacacataa
 25261 cggactatgtt tgatgtttt tgatgtttt tttttttttt tcaatgttgc cacaccc
 25321 aaaaatggc aatgttcttc ctgtcaaggac tggccatgttgc tttttttttt tttttttt
 25381 agccctcaat ctttttttttggatggccatgttgc tttttttttt tttttttttt
 25441 cgttccaaaaa ataaatggcc tcaataaaaatggccatgttgc tttttttttt
 25501 gtccatgttgc aattttacttgc tgctttttttt tttttttttt
 25561 tgccatgttgc gaggccatgttgc tttttttttt tttttttttt
 25621 caacccatgtt gaaatattttt tgatgtttgc tttttttttt
 25681 attttttttt gttttttttt tttttttttt
 25741 aacatataaa atgttgcacat ataaatgttgc tttttttttt
 25801 aaaaatggccaa gaaatgttgc aatattttttt
 25861 agactatgttgc tggttgc tttttttttt
 25921 aattttacttgc tttttttttt
 25981 agacccatgtt gttttttttt
 26041 aatgttgcattt tttttttttt
 26101 aatgttgcattt tttttttttt
 26161 tttttttttt
 26221 tttttttttt
 26281 tttttttttt
 26341 tttttttttt
 26401 tttttttttt
 26461 tttttttttt
 26521 tttttttttt
 26581 tttttttttt
 26641 tttttttttt
 26701 tttttttttt
 26761 tttttttttt
 26821 tttttttttt

FIG. 10 Com't

18 / 106

FIG. 10 Con't

19/106

1 - ATATTTAGGTTTAACTTACCCRGAAAAGCCAAACRACCTCGATCTCTGTAGATCTGTT - 60
 - I L G F Y L P R K S Q P T S I S C R S V
 - Y * V F T Y P G K A N Q P R S L V D D F
 - I R F L P T Q E K P T N L D L L * I C S
 61 - CTCCTAAACCAACTTAAATCCTGCTGACTGGCTCGCTCGCTGCATGCCATTAGTCACCTAC - 120
 - L * T N F K I C V A V A R L H A * C T Y
 - S K R T L K S V * L S L G C M P S A P T
 - L N E L * N L C S C R S A R A C L V H L R
 121 - GCAGTATAAACATAATAATTITACTGTCGTGACAAGAAACGAGTAACCTGTCCTCT - 180
 - A V * T I I N F T V V D K K R V T R P S
 - Q Y K Q * * I L L S L T R N E * L V P L
 - S I N C R * C R * Q E T S N S S L F
 181 - TCTGCAGACGCTACGGTTTCTCGCGTGTGCACTGCACTACCATACCTAGTTTC - 240
 - S A D C L R F R P C C S R S S A Y L G F
 - L Q T A Y G F V R V A V D H Q H T * V S
 - C R L L T V S S V Q S I I S I P R F R
 241 - GTCCGGGTGTGCCCCAAGGCTAGATGGAGGAGCTTGTCTTGGTGCAAGAGAAACAA - 300
 - V R V * P K G K N E S L V L G V N E K T
 - S G C D R K V R W R A L F L V S T R K H
 - P G V T E R * D G E P C C S W C Q R E N T
 301 - CACGTCRACTCAGTTGCCCTGCTCCAGGTAGAGACCTGCTAGTGCCGTGGCTTCGGG - 360
 - H V Q L L S P V L Q V R D V L V R G F G
 - T S N S V C L S P F R L E T C * C V A S G
 - R P T Q F Q A C P S G * R R A S A W L R G
 361 - GACTCTGGAAAGGGCCCTATGGAGGGCACGTGAACACCTCAAATGGCACTGGGT - 420
 - D S V E E A L S E A R E H L K N G T C G
 - T L W K R P Y R R H V N T S K M A L V V
 - L C G R G P I G G T * T P Q K W H L W S
 421 - CTAGTAGRCGCTGGAAAGGCGTACTGCCCCAGCTTGAACAGCCCCTATGGTTCATAAA - 480
 - L V E L E K G V L F Q L E O P Y V F I K
 - * * S W K K A Y C P S L N S P M C S L N
 - S R A G K R R T A P A * T A L C V H * T
 481 - CGTCTGGCTTAAGCACCATACTGGCCACAGGTGCTTGAAGCTGGTTGAGAAATG - 540
 - R S D A L S T N H G H K V V E L V A E M
 - V L M P * A P I T A T R S L S W L Q K N
 - F * C L K H Q S R P Q G R * A G C R N G
 541 - GACGGCATTCAGTAGCGTACGGCTATAACACTGGAGTACTCTGCCACATGGGC - 600
 - D G I Q Y G R S G I T L G V L V P H V G
 - T A F S T V V V A V * H W E Y S C H M W A
 - R H S V R * R Y N T G S T R A T C G R
 601 - GAAACCCCATTCGCACTACCGCATGTCCTCTCTGCAAGAACGCTATAAAGGAGCCG - 660
 - E T P I A Y R N V L L R K N G N K G A G
 - K P Q L H T A M F F F V R T V I R E P V
 - N P N C I P Q C S S S * E R * * G S R W
 661 - GGTCATCATGCTATGGCATGATGCTAAAGCTTATGACTTAAGGTGACCGACTGGCACTGAT - 720
 - G H S Y G I D L S Y D L G D E L G T D
 - V I A M A S I * S L M T * V T S L A L I
 - S * L W H R S K V L * L R * R A W H * S
 721 - CCCATTGGAAAGATTATGACMACAAACTGGNACACTAACGATGGCGATGGCACTCCGGTGA - 780
 - P I E D Y E Q N W N T K H G S G A L R E
 - F L K I M N K T G T L S M A V V H S V N
 - H * R L * T K L E H * A W Q W C T P * T
 781 - CTCACACTGAGCTCAATGGAGGTGCACTGCTCATGTCGACARACAAATTCTGTGGC - 840
 - L T R E L N G G A V T R Y V V D N N F C G
 - S L V S S M E V Q S L A M S T T I S V A
 - H S * A Q W R C S H S L C R Q Q F L W P

FIG. 11

20/106

841 - CCGAGATGGGTACCCCTTGTGATCAAAGATTTCGCCACGCCGGGCAAGTCATG - 900
 - P D G Y P L D C I K D F L A R A G K S M
 - Q M G T L L I A S K I F S H A R A S Q C
 - R W V P S * L H Q R F S S R T R G Q V N V
 901 - TGACTCTTCCGAAACACTTGATTACATCGAGTCGAAGAGGGTGTCTACTGCTGGCGT - 960
 - C T L S E Q L D Y I E S K R G V Y C C R
 - A L F P P N N L I T S S R R E V S T A A V
 - H S F R T T * L H R V E E R C L L L P *
 961 - GACCATGAGCATGAAATTGCCCTGGTCACTGAGCCTCTGATAAGRGCTACGAGCACAG - 1020
 - D H E H E I A W F T E R S D K S Y E R Q
 - T M S M K L P G S L S A L I R A T S T R
 - P * N O N C L V H * A L * * E L R A P D
 1021 - ACACACCTTCCGAAATAAGACTGCCAGAAAATTGACACTTCAAAGGGAAATGCCAAAG - 1080
 - T P F E I K S A K K F D T F K G E C P K
 - H P S K L R V P R N L T L S K G N A Q S
 - T L R N * E C Q E I * H F Q R G N P K V
 1081 - TTGTTGTTCTCTTAACTAACAAAGTCAGACCTGATCACCACGTGTTGAAAGAAG - 1140
 - F V F P L N S K V K V I Q P R V E K R K
 - L C F L L T Q K S K S E N H V L K R K R
 - C V S S * L K S Q S H S T T C * K E K D
 1141 - ACTGAGGTTTCTATGGGCGTAGCGCTCCTGACCTGTTGATCTCCACAGGGAGGT - 1200
 - T E G F M G R I R S V Y P V A S P Q E C
 - L R V S W G V Y Y A L C T L L H L H R S V
 - * G F H G A Y T L C V P C C I S T G V *
 1201 - AACATATGCACATGTTCTACCTGTGAAATGTANTCATGGCGTAGAAGTTCTGGCG - 1260
 - N N M H L S T L M K C N H C D E V S W Q
 - T I C T C L P * * N V I I A M K F H G R
 - Q Y A L V Y L D E M * S L R * S F M A D
 1261 - ACGTGGACTTCTGAAGCCACTTGTGACACTGTGGCACTGAAAATTAGTTATGAA - 1320
 - T C D F L K A T C E H C G T E N L V I E
 - R A T T F * K P L V N I V A L K I * L L K
 - V R L S E S H L * T L W H * K F S Y * R
 1321 - GGACCTACTACATGIGGTACTACCTACTATGTGTAGTGAARATGCCATGTCTGCC - 1380
 - G P T T C G Y L P T N T A V V V K M P C P A
 - D L L H V G T Y L L N L * * K C H V L P
 - T Y M W W V P T Y * C C S E N A M S C L
 1381 - TGTCAANGCCCAGAGATTGGCAGCTGAGCTAGTGTGAGATTATCACACCACTCAAC - 1440
 - C Q D P E I G P E H S V A D Y H N H S N
 - V K T Q R L D L S I V L Q I I T T T Q T
 - S R P R D W T * A * C C R L S Q P L K H
 1441 - ATTGAAACTGACTGCTCAAGGGAGGTAGGACTAGATGTTGGAGCTGTGTTGCC - 1500
 - I E T R L R K G G R T R C F G G C V F A
 - L K L D S A R E V G L D V L E A V C L P
 - * N S T P Q G R * D * M N F W R L C V C L
 1501 - TATGTTGGCTGCTATATATAGCTGCGTACTGGGTCTCTGCTAGTGTGATATGGC - 1560
 - Y V G C Y N K R A Y W V P R A S A D I G
 - M L L A A I I S V P T G F L V L V L I L A
 - C W L L * * A C L L G S S C * C * Y W L
 1561 - TCAGGCCATACTGGCATTACTGGTACATGGACATGGGAGGCTTGAATGGAGATCTCTTGAG - 1620
 - S G H T G I T G D N V E T L N E D D L E
 - Q A I L A L L V T M W R P * M R K I S L R
 - R P Y W H Y W * Q C G D L E * G S P * D
 1621 - ATACTGAGTCGTGAGCTGTTACATTAACATGGCTGGCGATTTCAATGAGTARGAG - 1680
 - I L S R E R V N N I N I V G D F H L N E E
 - Y * V V N V L T L T L L A I F I * M K R
 - T E S * T C * H * H C W R F S F E * R G

FIG. 11 Con't

21/106

1681 - GTTGCCATCATTTGGCATTCTTCGCTTCTACAAAGTGCCTTATTGRCACTATAAG - 1740
 - V A I I L A S F S R A S T S A F I D T I K
 - L P S F W H L S L L L Q V P L L T D * R
 - C H H F G I F L C F Y K C L Y * H Y K E
 1741 - AGTCTGATTAACAGTCCTTCANAAACCATGTGAGTCCTGCCTGAACCTATAAAGTACCC - 1800
 - S L D Y K S F K T I V E S C G N Y K V T
 - V L I T S L S K P L L S P A V T I K L P
 - S * L Q V F Q N H C * V L R * L * S Y Q
 1801 - AAGGGUANGCCCGTAAAGGTGCTGGACATTTGGACACAGAGATCAGTTAACACCA - 1860
 - K G K P V K G A W N I G Q Q R S V L T P
 - R E S P * K V L G T L D N R D Q F * R H
 - G K A R K R C L E H T T E I S F N T T
 1861 - CTGTTGGCTTTCCTCACAGGCTGCTGGTATTCAGATCAATTTCGCCCCACACTT - 1920
 - L C G F P S Q O A R G V I R S I F A R T L
 - C V V F P H R L L V L S D Q F L R A H L
 - U W S F L T G C W C Y Q I N F C A H T *
 1921 - GATGAGCAAACCACTCAATTCTGATTTCAGAACAGGACAGTCACCAACTGTGATGGT - 1980
 - D A R N H S I P D L Q R A A V T I L D G
 - M Q Q T T Q F L I C K E Q L S P Y L M V
 - C S K P L N S * F A K S S C H T * W Y
 1981 - ATTTCGACAGTCATACGCTTGTGACGCCATGTTTATACCTTCAGACCTGCTCACCC - 2040
 - I S E Q S O L R L V D A M V Y T S D L L T
 - F L N S H Y V L S T P W F I L Q T C S P
 - F * T V I T S C R R H G L Y F R P A H Q
 2041 - AACAGTGTCTTATTATGCGCATATGTRACAGTCGGCTGTGACACAGACTCTCAGTGG - 2100
 - N S V I I M A Y T V G G L V Q Q T S Q W
 - T V S S L L W H M * L V V D L Y N R L L S G
 - Q C H Y G I M C N A W W N S C T T D F S V V
 2101 - TTGCTTAATCTTTGGCACACGTTGAAAAACTCAGGCCATCTTGAATGGATTGAG - 2160
 - L S N L L G T T V E K L R P I F E W I E
 - C L I F W A L L L K N S G L S L N G L R
 - V * S F G H Y C * K T Q A Y L * M D * G
 2161 - GCGAAACATTAGTCAGGAGTTGATTCAGAGTCGGAGATCTCAARATTTC - 2220
 - A K L S S A G V E F L K D A W E I L K F L
 - R N L V Q E L L N F S R H M L G R F S N F S
 - E T * C R S * I S Q G C L G D S Q I S H
 2221 - ATTACAGGTGTTTIGACATGTCAGGGCAATTCAGGGTCTTCAGATAACATCAG - 2260
 - I T G V F D I V K G Q I Q V A S D N I K
 - L Q V F L T S S R V K Y R L L Q I T S R
 - Y R C F * H R G S N T G C F R * H Q G
 2281 - GATTGTTAAAGTCATTGATGTTGATTCAGAGGACTCGAAATGTCATGATCAA - 2340
 - D C V K C F I D V V N K A L E M C I D Q
 - I V * N A S L M L L T R H S K C A L I K
 - L C K M L H * C C * Q G T R N V H * S S
 2341 - GTCACTATCGCTGGCCAAAGTGGCATCAGTCAGGTGAGTCCTCGCTCGA - 2400
 - V T I A G A K L R S L N L G E V F I A Q
 - S L S L A Q S C D H S T * V K S S S S L K
 - H Y R W R K V A I T Q L R * S L D H R S K
 2401 - AGCAAGGAGACTTACCGTCAGTCATACGTTGGCAAGGAGCAGCTGCAACTCTCATGCC - 2460
 - S K G L Y R Q C I R G K E Q L Q L L M P
 - A R D F T V S V Y V A R S S C N Y S C L
 - Q G T L P S V Y T W O G A A A T T H R S
 2461 - CTTAAGGCACRANAGAAGTAACCTTCTGAGGTTGATTCAGATGACAGCTACTTAC - 2520
 - L K A P K E V T F L E G D S H D T V L T
 - L R H Q K K * P F L K V I H M T Q Y L P
 - * G T K R S N L S * R * F T * H S T Y L

FIG. 11 Con't

22/106

2521 - TCTGAGGAGGGTTCTCAAGAACGGTAACTCGAGACACTCGAGACGCCCGTTGATAGC - 2580
 - S E V V L K N G E L E A L E T P V D S
 - L R R R L F S R T V N S S K H S R R F L I A
 - * G G C S Q E R * T R S T R D A R * * L
 2581 - TCTCACAAATGGAGCTATCGTCGGCACACCGAGTCCTGTGAATGGCCCTATGCTCTTAGAG - 2640
 - F T N G A I V G T P V C V N G L M L L E
 - S Q M E L S S A H Q S V * M A S C S * R
 - H K W S Y R R H T S L C K W P H A L R D
 2641 - ATTAAGGACAAGAACATACTCGCAATTCTCTCTGGTTTACTGGCTACAAACAAATGTC - 2700
 - I K D K E Q Y C G A L S P G L L D A T N N V
 - L R T K N T N T A H C L L V Y W L Q T M S
 - * G C R T I L R I V S W F T G Y K Q C L
 2701 - TTTGCGTTAAAGGGTGCAGCCATTAAAGGTGTAACCTTTGGAGAAGATACTGGTT - 2760
 - F R L K G G A P I K G V * P L E K I L F G
 - F A * K G V H Q L K V * P L E K I L F G
 - S L K R G C T N * R C N L W R R Y C L G
 2761 - GAAGTICAGGGTTACAGAAATGTCAGAGATCAGCTTGAGCTTGATGACCTGTGACAA - 2820
 - E V Q G Y K N V P R I T F E L D E R V D K
 - K F K V T R M * E S H L S L M N V L T K
 - S R L Q E C E N H I * A * * T C * Q S
 2821 - GTGCCTTAAAGAAAGTGTCTGCTCATCTGTTGAGATCCGGTACCGAAGTTACTGAGTT - 2880
 - V L N E K C S V Y T V E S G T E V T E F
 - C L M K S A L S T L L N P V P K L L S L
 - A * * K V L C L H C * I R I Y R S Y * V C
 2881 - GCATGIGTGTGCGAGGGCTGTGTGRAGACTTACRACCGTTCTGATCTCCCTTAC - 2940
 - A C V V A E A V V V K T L Q P F V S D L L T
 - H V L * Q R L L * R L Y N Q F L I S L P
 - M C S R G C L * E D F T T S F * S P Y Q
 2941 - AACATGGTATTGATCTTGTAGTGGAGTGTAGCTACATTCTACTTATTGATGATGCT - 3000
 - N M G I D L D E W S S V A T F Y L F D D A
 - T N V U L I L M S G V * L H S T Y L M M L
 - H G Y * S * * V E C S Y I L L I * * C W
 3001 - GEGTGAAGAACATTTCATCATGTTGTTCTTACCCCTCGAGTGGAGAAGAA - 3060
 - G E E N F S S R M Y M C S F Y P P D E E E
 - V K K T F R H V C I V P F T L Q M R K K
 - * R K L F I T Y V L F L L P S R * G R R
 3061 - GAGGRGAGTGCAGAGTGTGGAGGAGAAGAAATTGTGAACCTGTGACATGAGTACCGT - 3120
 - E D D A E C E E E I D E T C E H E Y G
 - R T M Q S V R K K K L M K P V N M S T V
 - G R C R V * C V R * N L * T + V R Y
 3121 - ACAGAGATGATTATCAGGTCCTCCCTCTGGATTGTTGTCGCTCGAGCTGAACAGCTGA - 3180
 - T E D D Y Q G L P L E F G A S A E T V R
 - Q R M I I K V S L W N L V P Q L K Q F E
 - R G * L S R S P S G I W C L S * N S S S
 3181 - GTTGAGGAGAGAGAGAGAGAGACTGCGCTGGATGACTACTGAGCAGTCAAGAGTTGAG - 3240
 - V E E E E E E B D W L D D T T E Q S E I E
 - L R K K K R K T G W M I D L L S N Q R L S
 - * G R R R G R L A G * Y Y * A I R D * A
 3241 - CCAGAACAGAACCTACACCTGAGAACAGCTTAACTGAGTTTACTGGTTTAAACAT - 3300
 - P E P E P T P E E P V N Q F T G Y L K L
 - Q N Q N L H L K N Q L I S L L V I * N L
 - R T R T Y T * R T S * S V Y W L F K T Y
 3301 - ACTGACATGTTGCCATTAAATGTTGACATGTTAGGGCCACAAGTGTAAATCCT - 3360
 - T D N V A I K C V D I V K E A Q S A N P
 - L T M L P L N V L T S L R R H K V L I L
 - * Q C C H * M C * H R * G G T K C * S Y

FIG. 11 Con't

23/106

3361 - ATGGGTGATGTTAAATGCTGCTAACATACACCTGAAACATGGTGGGGTAGCAGGTGCA - 3420
 - M V I V N A A N I H L K H G G G V A G A
 - W * L * M L D T Y T * N M V V V V * Q V H
 - G D C K C C * H T P E T W W W C S R C T
 3421 - CTCACACAGGCCAACCATTGGTGCCTGCAAAGGGCAGTCATGATTACATTAAGCTAAT - 3480
 - L N K A T N G A M Q K E S S D D Y I K L N
 - S T R Q P M V P C K R R V M I T L S * M
 - Q Q G N Q W C H A K G E * * L H * A K W
 3481 - GGCCCTCTACAGTAGGGGGCTTGTGCTTCTGGACATAATCTGGTAAAGACTG - 3540
 - G P L T V G G S G C L D S G H N L A K K C
 - A A L L Q * E G L V C F L D I I L D R S V
 - P S Y S R V L F A F W T * S C * E V S
 3541 - CTGCGATGTTGGACCTAACCTAAATCGAGGTGGAGACATCACCGCTTAAAGCGCR - 3600
 - L H V V G P N L N A G E D I O L L K A A
 - C M L L D L T * M Q V R T S S F L R Q H
 - A C C W T * P K C R * G H P A S * G S I
 3601 - TAGGAAATTCTTCAATTCAAGGACATCTACTTGACCATTTGGTCAAGGATATT - 3660
 - Y E N F N S Q D I L L A P L L S A G I F
 - M K I S I S I H R T S Y L H H C C Q Q A Y L
 - * K F Q F T G H L T C T I V V S R H I W
 3661 - GGTGCTAAACCCATTCTGCTTAACTAACAGTGTGGTGCAGACGGTTCTGACACAGGTTTAT - 3720
 - G A K P L Q S L Q V C V Q T V R T Q V Y
 - V L N H F Q S L Y K C A C R R F V H R F I
 - C * T T S V F T S V R A D G S Y T G L Y
 3721 - ATTGCGATGCAATGACAAGCTCTTATGAGCGATTGTCATGGATTATCTGATAACCTG - 3780
 - I A V N D K A L Y E Q V V M D Y L D D N L
 - L Q S M T K L F M S R L S W I I L I T *
 - C S Q * Q S L * A G C H G L S * * P E
 3781 - AAGCCTAGGGAGGACCTTAAACAGGAGGCCAACACAGAAGATTCCAAAATC - 3840
 - K P R V E A P K Q O E E F P N T E D S K T
 - S L E W K H L N K R S H Q T Q K I P K L
 - A * S G S T * T R G A T K H R R F Q N *
 3841 - GAGGGAACTCTGCTGACAGGCTCTGCGATGTGARGGCCAAAATTAAAGGCTGCGATT - 3900
 - E E K S V V Q K P V D K P K I K A C I
 - R R N L S Y R S L S M * S Q K L R P A L
 - G E I C R T E A C R C E A K N * G L H *
 3901 - GATGAGGTACACACACTGGRAGAACTAAGTTCTTACCATAAAGTACTCTGTTT - 3960
 - D E V T T T L E E T K F E L T N K B L L F
 - M R L P Q H W K K L S F D P I S Y S C L
 - * G H N T G R N * V S Y Q * V T L V C
 3961 - GCTGATATCAATGTAAGCTTACCATGATTCTCGAACATGCTTGAAGGTGAAGATATG - 4020
 - A D I N G K L Y H D S Q N M L R G E D M
 - L I S M V S F T M I L R T C L E V K I C
 - * Y Q W * R L P * F S E H A * R * R Y V
 4021 - TCTTCCCTGAGAGGATGCACTTACATGGTAGGTATCAGTGTATCAGTGTATC - 4080
 - S F L E B K D A P Y M V G D V I T S G D I
 - L S L R R M H D T W * V M L S L V V I S
 - F P * E G C T L B G R * C Y H * W * Y H
 4081 - ACTTGTTGTTAATACCTCCAAAAAAGGCTGGTGGCAGTACTGAGATGCTCTCAAGAGCT - 4140
 - T C V V I P S K K A G G T T E M L S R A
 - L V L * Y P P K R L V A L L R C S Q E L
 - L C C N T L Q K G W W H Y * D A L K S F
 4141 - TTGAGAGAAGTGCCTGAGTATATAACCAAGCTACCTGGACARAGGATGTGCTGGT - 4200
 - L K K V P V D E Y I T T Y P G Q G C A G
 - * R K C Q L M S I * P R T L D K D V L V
 - E E S A S * * V Y N H V P W T R M C W L

FIG. 11 Con't

24/106

4201 - TATAACCTTGAGGAAGCTAACAGACTGCTCTTAAGAAATGCAATTCTGCATTTTATGTACTA - 4260
 - Y T L E E A K T A L L K C K S A F Y V L
 - I H L R K L R L L L R N A N A N L H F M Y Y
 - Y T * G S * D C S * E M Q I C I L C T T
 4261 - CCTTCAGAACGACCTTAATGCTTAAGGAAGAGATTCTAGGAACITGATCTGGATTTGAGA - 4320
 - P S E A P N A K E E I L G T V S W N L R
 - L Q K H L M L R K R F * E L Y P G I * E
 - F R S T * C * G R D S R N C I L E F B R
 4321 - GAAATGCCTGCTCATGCTGAAGAGCACAGAAAATTAATGCCATATGCTATGGATGTTAGA - 4380
 - E M L A H A E E T R E L M P I C M D V R
 - K C L L M L K R Q E N * C L Y A W M L E
 - N A C R A C C * R D K K I N A Y M H G C * S
 4381 - GGCATAATGGCACCATCCACGTAAGTATAAGGAATTAAATTCAGAGGGCATCGT - 4440
 - A I M A T I Q R K Y K G I K I Q E G I V
 - P * W Q P S N V S I K E L L K P K R A S L
 - H N G N H P T * V * R N * N S R G H R *
 4441 - GACTAAGGGTCTCCGATTCTCTTTTATCTAGTAAAGAGCTGTAGCTTCTATTATTACG - 4500
 - D Y G V R F F Y T S T K E P V A S I T
 - T M V S D S S F I L V K S L * L L L L R
 - L N C P I I L L Y * * R A C S F Y Y Y E
 4501 - ARGGCTGAACCTCTRAATGAGCCGCTGTCAACATGCCATTGGTTATGTGACACATGGT - 4560
 - K L N S L N E P L V T M P I G Y V T H G
 - S * T L * M S R L S Q C Q L V M * R M V
 - A E L S K * A A C H N A N W L C D T W F
 4561 - TTATACCTGAGAGGCTCGCCGGCTGATGGCTTCTCTAAAGCTCTGGCTAGTGTCA - 4620
 - F N L E E A A R C M R S S L K A P A V V S
 - L I L K R L R A V C V L L K L L P * C Q
 - * S * R G C A L Y A S * S S C R S V S
 4621 - GTATCATCACCAGCTGCTTACTACATATAATGATACCTCTACTTGCTCATCAAAGACA - 4680
 - V S S P D A V T T Y N G Y L T S S S K T
 - Y H B Q M D L L H I M D T S L R H Q R H
 - I I T R C C Y Y I * W I P H F V I K D I
 4681 - TCTGAGGAGCCTTGTAGAACAGCTTCTGGCTGGCTTACAGAGATTGGTCTTAT - 4740
 - S E E H F V E T V S L A G S Y R D W S Y
 - L R S T L * K Q F L W L A L T E I G P I
 - * G A L C R N S F F G W L L Q R L V L F
 4741 - TCAGGACACGOTACAGAGTAGGGTGTGAMTTCTTAAGCTGGTGACAATTGTTAC - 4800
 - S G Q R T E L G V E F L K R G D K I V Y
 - Q D S V Q S * V L N F L L S V V T K L C T
 - R T A Y R V R I S * I S * A W * Q N C V P
 4801 - CACACTCTGGAGCCCCCTGGAGTTCTGTAGCGGTGAGGTCTTCTGACAAA - 4860
 - H T L E S P V S F I L T V R F F H L T N
 - H S G R E P R R V S S * R * G S F T * Q T
 4861 - CTAAGAGGCTCTTATCCCTGGGGAGGTTAACAGACTATAAAGCTTACACACTGTTAC - 4920
 - L K S L S L R E V K T I K V F T T V D
 - * R V S Y P C G R L R L * K C S Q L W T
 - K E S L I P A G G * D Y K S V H N C G Q
 4921 - AACACTAATCTCCACACACAGCTTGTGGATATCTATGACATATGGACAGCAGTTGGT - 4980
 - N T N L H T Q L V D M S M T Y G Q Q F G
 - T L I S T H S L W I C L * H M D S S L V
 - H * S P H T A C G Y V Y D I W T A V W S
 4981 - CCCACATACTGGATGGTCTGTGTTACAAAATTAAACCTCATGTAATCATGAGGT - 5040
 - P T Y L D G A D V T K I K P H V N H E G
 - Q H T W M V L M L Q K L N L M * I M R V
 - N I L G W C * C Y K N * T S C K S * G *

FIG. 11 Con't

25/106

5041 - AAGACTTTCTTGTAACCTAGTGAATGACACACTACGTAGTAGAAGCTTCGAGTACTAC - 5100
 - K T F F V L P S D D T I R S E A F E Y Y
 - R L S L Y Y L V M T H Y V V K L S S T T
 - D F L C T T * * * H T T * * S F R V L P
 5101 - CATACTCTGATGAGACTTCTTGATGGTAGTACATGCTGCTTTAACCCACACAAAGAAA - 5160
 - H T L D E S F L G R Y M S A L N H T K K
 - I L L M R V F L V G T C L L * T T Q R N
 - Y S * * E F S W * V H V C F K P H K E M
 5161 - TGGAATTCTCAAGATTGGTGTTRAACCTCAATTAAATGGGCTGATAACAATGTTAT - 5220
 - W K F F P Q V G G L T S I K W A D N N N C Y
 - G N F L K L L V V * L Q L Q N G L I T I V I
 - E I S S S W W F N * M G * * Q L L F
 5221 - TTGCTCTAGTTTATAGCACTCAGCCTGAGACTCAATTCAATGACCGGACT - 5280
 - L S S V L L A L Q Q L E B V K F N A P A L
 - C L V F Y * H F N S L S K N S M H Q H F
 - V * C F I S T S T A * S Q I O C T S T S
 5281 - CAAGAGSCTTATTATAGGCCGCTCGTGTGATGCTGCTAACCTTTGGACTCATACTC - 5340
 - Q E A Y Y R A R A G D A A N F C A L I L
 - K R L I I P V L V M M L L T F V H S Y S
 - R G L L * S P C W * C C * L L C T H T R
 5341 - GCTTACAGTAAATAAAACTGCGAGCTTGGTGTGATGTCAGAGAAACTATGCCCATTT - 5400
 - A Y S N K T V G E L G D V R E T M T H L
 - L T V N K L L A S L V M S E K L * P I F
 - L Q * * N C W R A W * C Q R N Y D P S S
 5401 - CTACAGCATGCTAATTGGAAATCTGCAAAGCGAGCTCTAATGGGTGTAAACATGT - 5460
 - L Q H A N L E S A K R V L N V V C K H C
 - Y S M L I W N L Q S E F L M W C V N I V
 - T A C * F G I C K C A S * C G V * T L W
 5461 - GGTCAAGAAACTACTACCTTAACGGCTGTAGAGCTGTGATGTTATGGTACTCTATCT - 5520
 - G Q K T T T L T G V E A V M Y M G T L S
 - V R K L L L P * R V * K L * C I W V L Y L
 - S E N Y Y L N G C R S C D V Y G Y S I L
 5521 - TATGATRATCTAACAGCAGCTTCCCATTCATGTGTGTGTGTGTGTGTACCAA - 5580
 - Y D N L K T G V S I P C V C G R D A T Q
 - M I I L R Q V F P F H V C V V V V M L H N
 - * * S * D R C F H S M C V W S * C Y T I
 5581 - TATCTRGTACACAAAGACTCTTGTATGAGCTGTGCTGCACACCTGCTGATGATAAA - 5640
 - Y L V Q Q E S S F V M M S A P P A E Y K
 - I * Y N K S L L L L * C L H H L L S I N
 - S S T R V F F C V D V C T T C * V * I
 5641 - TTACAGCAAGTACATCTATGCGGATGAGTCACTGTAACATACGTGTGGTCAAT - 5700
 - L Q Q G T F L C A N E Y T G N Y Q C G H
 - Y S K V H S Y V R M S T L V T I S V V I
 - T A R Y I L M C E * V H W * L S V W S L
 5701 - TAACRCTCATACCTGCTTGGAGACCCCTATCCTATGAGGGCTCACCTTACAAAG - 5760
 - Y T H I T A K E T L Y R I D G A H L T K
 - T D I * L L R R P S I V L T E L T L Q R
 - H S Y N C * G D P L S Y * R S S P Y K D
 5761 - ATGTCAGGATCAAAGGACCTGAGCTGATGGTCTACAGGAAACATCTACACTAC - 5820
 - M S E Y K G P V T D V F Y K E T S Y T T
 - C Q S T K D Q * L M F S T R K H L T L Q
 - V R V Q R T S D * C F L Q G N I L H Y N
 5821 - ACCATCAAGCCCTGTGTGATMAACTCGATGGAGTACTTACACAGAGATTGACCAA - 5880
 - T I K P V S Y K L D G V T Y T E I E P K
 - P S S L C R I N S M E L L T Q R L N Q N
 - H Q A C V V * T R W S Y L H R D * T K I

FIG. 11 Con't

26/106

5881 - TTGGATGGGTATAAAGGATAATGCTTACTATACAGAGCAGCCTATAGACCTTGT A - 5940
 - L D G Y Y K K D N A Y Y T E Q P I D L V
 - W M G I I K R I M L T I Q S S L * T D Y
 - G W L * K G * C L L Y R A A Y R P C T
 5941 - CCAACTCAACCCATTACCAAMTGGAGTTTGATAATTTCAAACTCACATGTTCAAACCA - 6000
 - P T Q P I F N A S F D N F K L T C S N T
 - Q L N H Y Q M R V L I I S N S H V L T Q
 - N S T I T K C E F * F Q T H M F * H K
 6001 - AAATTTCGTATGATTTAACATGACAGGCTTCACAAAGGCCAGCTCACAGAGCTA - 6060
 - K F A D D D L N Q M T G F T K P A S R E L
 - N L L M I * I K * Q A S Q S Q L H E S Y
 - I C * F K S N D R L H K A S F T R A I
 6061 - TCTGTCACATTCTCCAGACTTGTGATGCGGAGTAGTGGCTATTGACTATAGACACTAT - 6120
 - S V T F F P D L N G D V V A I D Y R H Y
 - L S H S S Q T * M A M * W L L T I D T I
 - C H I L P R L E W R C S G Y * L * T L F
 6121 - TCAGCGGTTCAAGAAGGTGCTAAATTACTGCATRAGCCAATTTGGCACATTAAC - 6180
 - S A S F K K G A K L L H K P I V W H I N
 - Q R V S R K V L N Y C I S Q L F G T L T
 - S E F Q E R C * I T A * A N C L A H * P
 6181 - CAGGCTACAAACGAAACACGCTCAACCCAAACACTTGGGTTTACGTGTGCTTGGAGT - 6240
 - Q A T T K T F K P N T W C L R C L W S
 - R L Q P R Q R S N Q T L G V Y V V F G V
 - G Y N Q D N V Q T N K H L V F T L S L E Y
 6241 - ACRAAGCCAGTAGATCATCTAACATTCATTTGAAGTTCTGGCAGTAGAACACACAAAGA - 6300
 - T K P V D T S N S F E V L A V E D T Q G
 - Q S Q * I L Q I H L K F W Q * K T H K E
 - K A S R Y F K F I * S S G S R R H T R N
 6301 - ATGGACATCTGGCTTGTGAGACTAACAAACCCCCTGAAGAACATGAGTGGAAATCCT - 6360
 - M D N L R A C E S O Q P T S E E V V E N P
 - W T I L L V K V N N P P L K K * W K I L
 - G Q S C L * K S T T H L * R S S G K S Y
 6361 - ACCATACAGAAGGAAGTCATAGACTGACTGAARACTACCGAAGTGTAGGCATATGTC - 6420
 - T I Q K E V I E C D V K T T E V V G N V
 - P Y R R K S * S V T * K L P K L * A M S
 - H T E G S H R V * R E N Y R S C R Q C H
 6421 - ATACTAACCATCAGATGAGGTAAACTAACCCAGAGTTAGGGCATGAGATCTT - 6480
 - I L K P S D E G V K V T Q E L G R E D L
 - Y L N P H Q M K V L D K * H K S * V M R I L
 - T * T I R * R C * S N T R V R S * G S Y
 6481 - ATGGCGCTTATGTTGGAAACACRACGCTTACATTAAAGAAAACCTTATGCTTCTCACTA - 6540
 - M A A Y V E N T S I T I K K P N E L S L
 - W L L M W K T Q A L P F R L N L M S F H *
 - G C L C G K H K H Y H * E T * * A F T S
 6541 - GCCTTAGGTAAAACANTGCACTCATGTTAGCTGCAATTAATGCTGCAATTAAATGTTCTGG - 6600
 - A L G L K T I A T H G I A A I N S V P W
 - P * V * K Q L P L M V L L Q L I V F L G
 - L R F K N N C H S W Y C C N * C S L E
 6601 - AGTAAATAATTGCTTATGCTTACAAACCCATTCTGAGGACAGCAGCAATTACACATCAAT - 6660
 - S K I L A Y V K P F L G Q A A I T T S N
 - V K F W L M S N H S * D K Q Q L Q H Q I
 - * N P G L C Q T I L R T S S N Y N I K L
 - TGCCTAGAGAGATTAGCACACAGTGTGTTAACMATTATATGCTTATGTTACATTA - 6720
 - C A K R L A Q R V F N N Y M P Y V F T L
 - A L R D * H N V C L T I I C L M C L H Y
 - R * E I S T T C V * Q L Y A L C V Y I I

FIG. 11 Con't

27/106

6721 - TTGTTCCAATGTGTACTTTACTAAGTACCRATTCTAGAATTAGAGCTTCACCT - 6780
 - L F Q L C T F T K S T N S R I R A S L P
 - C S N C V U L L L K V P I L L E L H Y L
 - V P I V Y F Y * K Y Q F * N * S F T T Y
 6781 - ACRATATGTCTAAARATAGTGTTRAGACTGTCTCTAATTATGTTGGATGCCGCATT - 6840
 - T T I A K N S V K S V A K L C L D A G I
 - Q L L L K I V L R V L L N Y V W M P A L
 - N Y C * K * C * E C C * I M F G C R H *
 6841 - AATZATGTGAAGTCACCCAAATTCTAATATGTCTACACATCGTATGGCTATTGTG - 6900
 - N Y V K S P K F S K L E T I A M W L L L
 - I M * S H P N F L N C S Q S L C G Y C C
 - L C E V T Q I F * I V H N S L Y V A I V V
 6901 - TTARGTAATTGCTTAGGTCTCTAATCTGTGTACTGTCTTTGGTACTCTTATCT - 6960
 - L S I C C L G S L I C V T A R F G V L L S
 - * V F A * V L * S V * L L L L V Y S Y L
 - K Y L L R F S N L C N C C F W C T L I *
 6961 - AATTTTGTCGCTCTCTTATGTAAATGGCCTAGAGAATTGTATCTAAATGGCTTAC - 7020
 - F N G F A P S E Y C N G V R E L Y L N S S N
 - I L V L L L I V M A L E N C I L I R L T
 - F N C S F S L L * W R * R I V S * F V * R
 7021 - GTTACTACTATGGATTCTGTGAGGTTCTTCCTCTGAGCATTTGGTTAATGGATA - 7080
 - V T T M D F C E G S F F P C S I C L S G L
 - L L W I S V K V L F L R A A F V * V D *
 - Y Y G F L * R F F S S L Q H L F K W I R
 7081 - GACTCCCTTGATCTTATCGCCTCTGAAACCATTCAGGTGAGGATTTATCGTACAG - 7140
 - D S L D S Y P A L E T I Q V T I S S Y K
 - T P L I L I Q L L K P F R * R F F H R T S
 - L P * F L S S S * N H S G D D F I V Q A
 7141 - CTAGACTTGACATTTTAGCTCTGGCTCTGGGTTTTGGCATPATATGTTGTCACA - 7200
 - L D L T I D G L A A E W V L A Y L F T
 - * T * Q F * V W P L S G F W H I C C S Q
 - R L D N F R S G R * V G F G I Y V V H K
 7201 - AAATCTTATTTATAGCTCTTCAGCTATAATGCTAGGTGTTCTTGGCTATTTGCT - 7260
 - K F F Y L L G L S A I M Q V F F G Y F A
 - N S F I Y * V F Q L * C R C S L A I L L
 - I L L F I R S F S Y N A G V L W L F C *
 7261 - AGTCATTCTCATCAGCAATTCTGGCTCATGGTTATCATAGTATTGTACAAATGGC - 7320
 - S H F I S N S W L M W F I I S I V Q M A
 - V I S S A I L G S C G L S L V L Y K W H
 - S F H Q F L A H V V V H * Y C T N G T
 7321 - CCCSTTCTGCAATGGTAGGTACATCTCTCTGGCTCTTCTACTACRATATGGAG - 7380
 - P V S A M V R M Y I F F A S F Y Y I W K
 - P F L Q W L G C T S S L L S T T Y G R
 - R F C C N G * D V H H L C F F L L H M E E
 7381 - AGCTATGTCATCATGGATGGCTACCTCTGGACTCTGGATGTGTCTATAGGGC - 7440
 - S Y V H I M D G C T S S T C M M C Y K R
 - A M F I S W M V A P L R L A * C A I S A
 - L C S Y H G W L H L F D L H D V L * A Q
 7441 - AATCGTGCCRACCGCTTGGTAGTACAATCTGTATGGCATGAAGAGATCTTCTAT - 7500
 - N R A T R V E C T T I V N G M K R S F Y
 - I V P H A L S V Q L L L M A * R D L S M
 - S C H T R * V Y N Y C * W H E E I F L C
 7501 - GTCTATGCAATGGAGCCGGTGCCTCTGCAAGACATCACAATTGGAAATGCTCAATTGT - 7560
 - V Y A N G G R G F C K T H N W N C L N C
 - S M Q M E A V A S A R L T I G I V S I V
 - L C K W R P W L L Q D S Q L E L S Q L *

FIG. 11 Con't

28/106

7561 - GACACATTTGCACTGGTAGTACATTCAATTAGTGATGAAGTTGCTCGTAGTTGCACTC - 7620
 - D T F C T G S T F I S D E V A R D L S L
 - T H F A L V V H S L V M K L L V I C H S
 - H I L H W * Y I H * * S C S * F V T P
 7621 - CAGTTTAAGGACCAATCACCCCTACTGACAGCTACATCGTATATGTTGATAGTGTCT - 7680
 - Q F K R P I N P T D Q S S Y I V D S V A
 - S L K D Q S T L L T S H R I L L I V L L
 - V * K T N Q P Y * P V I V Y C * C C C
 7681 - GTGAAAGATGGCGCCTCACCTCTACTTGTGACAAGGCTGGTCAAAGACCTATGAGAGA - 7740
 - V K N G A L H L Y F D K A G Q K T Y E R
 - * K M A R F T S T L T R L V K R P M R D
 - E K W R A S P L * Q G W S K D L * E T
 7741 - CATCGCTCTCCATTTGCAATTAGACAATTGAGACCTAACAACTAAAGGTCA - 7800
 - H P L S H F V N L D N L R A N N T K G S
 - I R S P I L S I * T I * E L T T L K V H
 - S A L P F C Q F R Q F E S * Q H * R F T
 7801 - CTGCTTATATGTCAGTTGATGGCAAGTCACATGGCAGAGTCCTAG - 7860
 - L P I N V I V F D G K S E C D E S A S K
 - C L L M S * F L M A S P N A T S L L L S
 - A Y * C H S F * W Q V Q M R R V C F * V
 7861 - TCTGCTCTGTACTACAGTCACTGATGTCGCAACCTATTCTGCTGACCAAGCT - 7920
 - S A S V Y S O L M C O P I L L L D Q A
 - L L L C T T V S * C A N L F C C L T K L
 - C F C V D Q S A D V F T Y S V A * P S S
 7921 - CTTGTCRACAACGTTGAGAGATGACTGAGATTTCTGATGATCTTGTCTAGTC - 7980
 - L V S S N V G D S T E V S V K M F D A Y V
 - L Y Q T L E I V L K F F L R C L M L M S
 - C I K R W R * Y * S F R * D V * C L C R
 7981 - GRCACCTTCTACCAACTTACTGCTCTATCGAAAACCTTAAGGCACCTGTCRCA - 8040
 - D T F S A T F S V P M E K L K A L V A T
 - T P F F Q Q L L V F L W K N L R R L L L Q
 - H L F S N F * C S Y G K T * G T C C Y S
 8041 - GCTCACAGGAGTACCAAAAGGTTAGCTTGTGTTGCTCTTCTACATTGTC - 8100
 - A H S E L A K G V A L D G V L S T F V S
 - L T T A S * Q R V * L * M V S F L H S C Q
 - S Q R V S K G C S F R W C P F Y I R V S
 8101 - GCTGCCGACAAGGTGTTGATGACCGATGTTGACACAAGGATGTTAGATGTC - 8160
 - A A R Q G V V D T D V T K D V I E C L
 - L P D K V L L D I P M L T Q R M L L N V S
 - C P T R C C * Y R C * H K G C Y * M S Q
 8161 - AAACCTTCACATCACTGACTTAGAGACTGACAGGTGACAGTGTAAACATTCTGTC - 8220
 - K L S H H S D L E V T G D S C N N F M L
 - N F H I T L T * K * Q V T V V T I S C S
 - T F T S L * L R S D R * Q L * Q F H R H
 8221 - ACCCTTAAGGTTGAAACATGACCCCCAGAGATCTTGGCCATGATGACTGTAAT - 8280
 - T Y N K V E N M T P R D L G A C I D G N
 - P I I R L K T * R P E I L A H V L T V M
 - L * * G * K H D A Q R S W R M Y * L * C
 8281 - GCAAGGCATATCAATGCCCNACTAGCAAAAAGTCACACANTGTTCACTCATGGAAAGTA - 8340
 - A R H I N A Q V A K S H N V S L I W N V
 - Q G I S M P K * Q K V T M F H S S G M *
 - K A Y Q C P S S K K S Q C F T H L E C K
 8341 - AAAGACTACATGCTCTTATGCAACAGCTGCGTAAACAAATTGCTACTGCTGCCAGAG - 8400
 - K D Y M S L S E Q L R K Q I R T A A K K
 - K T T C L Y L N S C V N K F V L L P R R
 - R L H V F I * T A A * T N S Y C C Q E E

FIG. 11 Con't

29/106

8401 - AACRACATACCTTTACACTAATGTGCTCACRACATAGACAGGGTGTCAATGTCATRACT - 8460
 - N N I P F T L T C A T T R Q V V N V I T
 - T T Y L L R * D V L Q D D R L S M S * L
 - Q H T F Y T N L C Y N * T G C Q C H N Y
 8461 - ACTAAAATCTACTAACGGGTTAAAGATTGTAACTTGTTTTAAACTTATGGCTTAAG - 8520
 - T K I S L K G G K I V S T C F K L M L K
 - L K S H S R V V R L L V L V I N L C L R
 - * N L T O G W * D C * Y L F * T Y A * G
 8521 - GCCACATTATTGTGCGTCTTGTCTGGTTTATATCCTATGCCAGTAGCTACAC - 8580
 - A T L L C C V L A A L V C Y I V M P V H T
 - P H Y C A F L L H W F V I S L C Q Y I H
 - H I I V R S C C I G L L Y R Y A S T Y I
 8581 - TTGTCAAATCATGATGGTACACAAATGAAATCATGGTTAACAGCATTGGGTGT - 8640
 - L S I H D G Y T N E I I G Y K A I Q D G
 - C Q S M M V T Q N M K S L V T K P F R M V
 - V N P * N L H L K * N H W L Q S H S G W C
 8641 - GTCACTCGTACATCATTTCTACTGATGATTTGGCAATAAATACATGCTGGTTTGAC - 8700
 - V T R D I I S T D D C F A N K H A G F D
 - S L V T S F L L M I V L Q I N M L L V L T
 - H S * H H F Y * * L F C K * T C W F * R
 8701 - GCATGGTTTACGACGGTGGGTGCTAACAAATAATGACAAAAGCTGGCCCTGTAGTAGCT - 8760
 - A W F S Q R G G S Y K N D K S C P V V A
 - H G L A S V V V H T K M T K A R L * * L
 - M V * P A W F I Q K * Q K L E P C S S C
 8761 - GCTATCATACAGAGAGATTGTTTATAGTCCTGGCTTACCGGGTACTGTGCTGAGA - 8820
 - A I I T R E I G F I V P G L P G T V L R
 - L S L Q E R L V S * C L A Y R V L C * E
 - Y H K R D W F H S W A L T G Y C A E S
 8821 - GCAATCATGGTACTCTGCTATTCTACCCGGTTTTATGTGCGTGGCACAT - 8880
 - A I N G D F L H F L P R V F S A V G N I
 - Q S M V T S C I F Y L V F L L V L L A T F
 - N Q W * L L R F S T S C F * C C W Q H L
 8881 - TGCTACACCCCTCCAACTCATGGTACTGATGGTTTGTACCTCTGGCTTACCTGGCTTCT - 8940
 - C Y T P S K L I E Y S D F A T S A C V L
 - A T H L P N S L S I V I L L P L L A F L
 - L H T F Q T H * V * * F C Y L C L R S C
 8941 - GCTGCTGAGTGTACATTAAAGGTGCTATGGGCAACCTGTGCATATTGTTATGAC - 9000
 - A A E C T T I F K D A M G K P V F Y C Y D
 - L L S V Q F L R M L W A N L C H I V M T
 - C * V Y N F * G C Y Q G T C A I L L * R
 9001 - ACTAATTTGCTAGGGGTCTTCTCTTCTATGAGCTGTGGCTTACAGACACTGGTTATGT - 9060
 - T N L L E G S I S Y S E L R F P D T R Y V
 - L I C * R V L F L I V S F V Q T L V M C
 - * F A R G F Y F L * * A S S R H S L C A
 9061 - CTATGGATGGTTCCATACAGTTCTACRACATACCTGGAGGGTCTGTAGAGTA - 9120
 - L M D G S I I Q F P P N T Y L E G S V R V
 - L W M V P S Y S F L T L T W R V L L E *
 - Y G W F H H T V S * H L P G G F C * S S
 9121 - GTACACACTTTGCTAGCTGAGTACTGTAGACAGGTACATGCAANGGTAGAGGT - 9180
 - V T T F D A E Y C R H G T C E R S E V G
 - * Q L L M L S T V D M V H A K G Q K * V
 - N N F * C * V L * T W Y M R K V R S R Y
 9181 - ATTTGCCATCTACCACTGGTATGGGTTCTTATATGAGCATACAGGCTCTATCA - 9240
 - I C L S T S G R R W V L N N E H Y R A L S
 - F A Y L P V V D G F L I M S I T E L Y Q
 - L P I Y Q W * M G S * * * A L Q S S I R

FIG. 11 Con't

30/106

9241 - GGRGTTTCTGTGGTGTGATGCATGATCTCTAGCTAACATCTTACTCCTCTGTG - 9300
 - G V F C G V D A M N L I A N I F T P L V
 - B F S V V V L M R * I S * L T S L L L L C
 - S F L W C * C D E S H S * H L Y S S C A
 9301 - CAACCGTGGTGGTGCCTTTAGATGTGTCCTGCCTCACTAGTGGCTGGCTATTGCCATA - 9360
 - Q P V G A L D D V S A S V V A G G I I A I
 - N L W V L . * M C L D Q * W L V V L L P Y
 - T C G C F R C V C F S S G W W Y Y C H I
 9361 - TTGGTACTCTGTCGCCACTACTTTATGAAATCAGACGTGTTTTGAGTACAC - 9420
 - L V T C R A Y Y F K M F R R V F G E Y N
 - W * L V L P T T L * N S D V F V L V S T T
 - G D L C C L L Y E I Q T C F W * V Q P
 9421 - CATGGTGTGCTGCTAATGCACTTTGTTGTTGTTACTATACTCTGCTGGTA - 9480
 - H V V A A A N A L L F L M S F T I L C L V
 - M L L L D M H F C F * C L S L Y S V W Y
 - C C C * C T F V F D V F H Y T L S G T
 9481 - CCAGCTTACAGCTTCTGCCGGGAGCTACTACTGACTCTTACTTGACTGACATCTAT - 9540
 - P A Y S F L F G V Y S V F Y L Y L T F Y
 - Q L T A F C R E S T Q S F T C T * H S I
 - S Q Q L S R G S L L S L L L V D L I L F
 9541 - TTCACCATATGATGTTCTACCTGGCTCACCTCAATGGTTGCCATGTTCTCCATT - 9600
 - F T N D V S F P L A H L Q W F A M F S P I
 - S P M M F H S W L T F N G L P C F L L L
 - H Q * C F I L G S P M V C H V F S Y C
 9601 - TGCGCTTCTGGATAACRCAACTATGATTCTGTTCTGAGGACTGGCATGG - 9660
 - V P F W I T A I Y V F C I S L K H C H W
 - C L F G * Q Q S M Y M S V F L * S T A I G
 - A F L D N S N L C I L Y F S E A L P L V
 9661 - TTCTTTAACACTATCTTAGGAARAGTCTGTTAATGGAGTTACATTAGTACCTTC - 9720
 - F F N Y N L R K R V M E N G V T F S T F
 - S L T T I L G K E S C L M E L H L V P F S
 - L * Q L S * E K S H V * W S Y I * Y L R
 9721 - GAGGAGGCTGGCTTGGTACCTTTTGTCAACAAAGGAATGACCTAAATTGGCTAGC - 9780
 - E E A A L C T F L L N K E M Y L K L R S
 - R R L C F V Y L F A R Q Q G N V P K I A * R
 - G G C F V Y L F A R Q Q G N V P K I A * R
 9781 - GAGACACTGTGCCACTACACAGTATAACAGGTATCTGCTCTATATAACAGTACAAG - 9840
 - E T L L P L T Q Y N N R Y L A L Y N K Y K
 - R H C C H L H S I T G I L L Y I T S T S
 - D T V A T Y T V * Q V S C S I * Q V Q V
 9841 - TATTCAGTGGAGGCTTACAGATACACAGCTATGCTGAGCTGGCTGGCACTTAGCA - 9900
 - Y F S G A L D T T S Y R E A A C C H L A
 - I S V E P * I L P A I V K Q L A A T * Q
 - F Q W S L R Y Y Q L S * S S L L P L S K
 9901 - AAGGCTTAATGACTTTAGCAACTCAGGTGCTGATGTTCTACCAACACAGACA - 9960
 - R A L N D F S N S G A D V L Y Q P P Q T
 - R L * M T L A T Q V L M F S T N H H R H
 - G S K * L * Q L R C * C S L P T T T D I
 9961 - TCANTCACTCTGCTGTTCTGCAGAGTGTGTTTGGAAATGGCATCCCGTCAGGCARA - 10020
 - S I T S A V L Q S G F R K M A F P S G K
 - Q S L L L F C R V V L G K W H S R Q A K
 - N H F C C S A E W F * E N G I P V R Q S
 10021 - GTGAAAGGTGCTGATGTAACAGTAACCTGTGGACTAACACTCTTAATGGATTGGTGG - 10080
 - V E G C M V Q V T C G T T T L N G L W L
 - L K G A W Y K * P V E L Q L L M D C G W
 - * R V H G T S N L W N Y N S * W I V V G

FIG. 11 Con't

31/106

10081 - GATGRCRACAGTAACTGTGTCGAAGACATGTCATTTGGCACAGCAGAAGACATGCTTAATCCT - 10140
 - D D T V Y C P R H V I C T A E D M L N P
 - M T Q Y T V Q D M S F A Q Q K T C L I L
 - * H S I L S K T C H L H S R R H A * S *
 10141 - AACATGAGAGATCTGCTATCGCRAATCCAAACCATAGCTTCTGTTCAAGGCTGGCAAT - 10200
 - N Y E D L L I R K S N H S L V Q A G N
 - T M K I C S F A N P T I A F L F R L A M
 - L * R S A H S Q I Q P * L S C S G W Q C
 10201 - GTTCAACTCGGCTTATGGCATTGCTATGCRAAATGTCCTGCTTAGGCTTAAAGTGTAT - 10260
 - V Q L R V I G H S M Q N C L D R L K V D
 - F N F V L L A I L C K I V C L G L K L I
 - S T S C Y W P Y A K L S A * A * S * Y
 10261 - ACTCTAACCTTAAGACACCCAGTATAATTCTCGTATCCAAACCTGGTAAACATTT - 10320
 - T S N P K T P K Y K F V R I Q P G Q T F
 - L L T L R H P S I N L S V S N L V K H F
 - F * P * D T Q S V * I C P Y P T W S N I F
 10321 - TCACTCTACCATGCTACATGGTCAACCATGCTTATGGCTTATCAGTGCCATGAGCT - 10380
 - S V L A C Y N G S P S G V Y Q C A M R P
 - Q F * H A T M V H H L V F I S V P * D L
 - S S M L Q W F T I W C L S V C H E T *
 10381 - AACATCACCATTAAGGTTCTCCCTAAATGGTCACTGGTACTGGTTAACATT - 10440
 - N H T I K G S F L N G S C G S V G F N I
 - I I P L K V L S L M D H V V V L V L T L
 - S Y H * R F F P * W I M W * C W F * H *
 10441 - GATATGATGCGTCTCTCTGCTATATGCTCATATGGCTTCTCAGTGCCATGAGCTACAC - 10500
 - D Y D C V S F C Y M H M E L P T G V H
 - I M I A C L S A I C I C I W S F Q Q E Y T
 - L * R V F L L Y A S Y G A S N R S T R
 10501 - GCTGGTACTGACTTAGAGGTAATTCTAATGGTCATTGTCAGACAGCAACTGCGAG - 10560
 - A G T D L E G K F Y G P F V D R Q T A Q
 - L V L T * K V N N S M V H H L L T D K L H R
 - W Y * L R R * I L W S I T C * Q T N C T G
 10561 - GCTGCAGGTACAGCACACACCATTAACATTAATGGCTTGGCATGGCTGTTGCTGCTT - 10620
 - A A G T D T T I T L D N V L A W L Y A A V
 - L Q V Q T Q P * H * M F W H G C M L L D
 - C R Y R H N H N I K C F G M A V C C C Y
 10621 - ATCAATGGTATGAGGTGTTCTTCTATAGATCCACCATTTGAATGACTTTAACCTT - 10680
 - I N G D R W F L N R F T T T L N D F N L
 - S M V I G G F L I D S P L L * M T L T L
 - Q W * V V S * I H A H Y F E * L * P C
 10681 - GTGGCAATGAAGTACAATGATGACCTTGGCACACAAAGTATGTCATATGGGACCT - 10740
 - V A M K Y N Y E P L T Q C D H V D I L G P
 - W Q * S T T M N L * H R I M L T Y W D L
 - G N E V Q L * T F D T R S C * R I G T S
 10741 - CTTCCTGCTCAAACAGGATTGGCGTCTAGATATGTCGCTTGAAGAGCTGCT - 10800
 - L S A O T G I A V L D M C A A L K E L L
 - F L L K Q E L P S * I C V C L L * K S C C
 - F C S N R N C R L R Y V C C F E R A A A
 10801 - CAGAATGGTATGAAIGTCGACTATCTTGGTAGGCACTATTTAGAAGATGAGTTACA - 10860
 - Q N G M N G R T I L G S T I L E D F T
 - R M V * M V V L S L V A L F * K M S L H
 - E W Y E R S V Y P W * H Y F R R * V Y T
 10861 - CCATTTGATGTTAGACAACTCTGCTTACCTTCCAAAGGTAAGTCAAGAAAATT - 10920
 - P F D V V R Q C S G V T F Q G K F K K I
 - H L M L L D N A L V L P S K V S S R K L
 - I * C C * T M L W C Y L P R * V Q E N C

FIG. 11 Con't

32/106

10921 - GTTAAAGGGCACTCATCATGGATGCTTTAACCTCTTGTGCATCACTATTGATCTTGTT - 10980
 - V K G T H W M D L T F L T S L L I L V
 - L R A L I I G C F * L S * H H Y * F D F
 - * G H S S L D A F N F L D I T I D S C S
 10981 - CAAACTACACAGTGGTCACGTGTTTCTTGTGTTACAGAGATGCTTCCTGCCATTACT - 11040
 - Q S T Q V S L F F V Y E N A F L P F T
 - K V H S G H C F S L F T R M L S C H L L
 - K Y T V V T V F L C L R E C F L A I Y S
 11041 - CTTGGTATTATGCCAATTGCTGCATGTGCTATGCTGCTGTGTAAGCATAAGCACCGATTC - 11100
 - L G I M M A I R A C A M L L V K H K H A F
 - L V L W Q P L L H V L C C L L S I S T H S
 - W Y G N C A C * A * A R I L
 11101 - TTGCTGCTGTTCTGCTACCTCTCTGCAACAGTGTGTTACTTAAATGGCTCATCTG - 11160
 - L C A C F L L P S L A T V A Y F N M V Y M
 - C A C F C Y L L W Q Q L T L I W N S T C
 - V L V S V T F S C N S C L L * Y G L H A
 11161 - CCTGCTAGCTGGTGTATGCTCATGACATGGCTGATGGCTGACACTAGCTGCT - 11220
 - P A S W V M R I N T A C T W L E L A D T S L S
 - L L A G * C V S * H G L N W L T L A C L
 - C * L G D A Y H D M A * I G * H * L V W
 11221 - GGTTATAAGGCTTAAGGATTTGGTGTATGATGCTCAGCTTGTGTTCTTATCTCATG - 11280
 - G Y R L K D C I V L C M Y R S A L V L I L M
 - V I G L R I V L C M Q L Q L * F C L F S *
 - L * A * G L C Y V C F S F S F A Y S H D
 11281 - ACAGCTCCGGCTGGTGTATGATGCTGAGCTGCTGAGCTGTGGACACTGATGATGATGCTT - 11340
 - T A R T V Y D D R A R R R V W T L M M N V I
 - Q L A L F M M M M L L D V F G H * * M S L
 - S S H C L * * C C * T C L D T D E C H Y
 11341 - ACACCTGGTTACAAGCTACTACTGTGTTAGTCAAGCTATTCTCAAGCTGGGCC - 11400
 - T L V Y K V Y G N A L D Q A I S M W A
 - H L F T K S T M V M L * I K L F P C G P
 - T C L Q S L L W * C F R R S S Y F H V G L
 11401 - TTAGTTTCTGTAACTCTAACTACTCTGGTGTGGTACGACTATCATGTTTGTGCT - 11460
 - L V I S P T S N Y S G V V T T I M F L A
 - * L F L * P L T I L V S L R L S C F * L
 - S Y F C N L * L F W C R Y D Y H V F S *
 11461 - AGAGCTATACTGTTCTGTGGTGTGATTACCCATTGGTTATTACTGGCRAACCC - 11520
 - R A I V F F V C V E Y Y P L L F I T G N T
 - E L * C L C V L S I T H C Y L L L A T P
 - S Y C V C * V L P I V I Y Y W Q H L
 11521 - TTACAGTGTATCATGCTGTTATGTTCTTAGCTATGTGGCTGCTGCTACTTGGC - 11580
 - L Q C I M L V Y C F L G Y C C C C Y F G
 - Y S V S C L F I V S * A I V A A A T L A
 - T V Y H A C L L F L R L L L L L L W P
 11581 - CTTCCTGTTTACTCACCGTTACTTCAGCTTACTCTGGTGTATTGACTACTGGTC - 11640
 - L F C C L L N R Y F R L T L G V Y D Y L V
 - F S V Y S T V T S G L L L V F M T T W S
 - F L F T Q P L L Q A Y S W C L * D L G L
 11641 - TCTACACAAAGAATTAGGTATAGAATCTCCAGGGCTTTGGCTCCTAAGAGTGTATT - 11700
 - S T Q E F R Y M N S Q G L L P K S S I
 - L H K N L G I * T P R G F C L L R V V L
 - Y T R I * V Y E L P G R F A S * E * Y *
 11701 - GATGCTTCAAGCTTAACATTAACTGTTGGGTTATGGAGGTAAACCTGTATCAGGTT - 11760
 - D A F K L N I K L L G I G G K P C I K V
 - M L S S L T L S C W V L E V N H V S R L
 - C F Q A * R * V V G Y W R * T M Y Q G C

FIG. 11 Con't

33/106

11761 - GCTACTGTAACAGTCTAAATGTCGACGGTAAAGTCGCACATCTGTGGTACTGCTCTCGGTT - 11820
 - A T V Q S K M S D V K C T S V V L L S V
 - L L Y S L K G L T * S A H L W Y C S R F
 - Y C T V * N V * R K V H I C G T A L G S
 11821 - CTCACACAACTTAGAGTCAGGCTCATCTCTAAATGTCGGCACATGTGGTACRAACTCCAC - 11880
 - L Q Q O L R V E S S S K L W A Q C V Q L H
 - F N N L E * S H L L N C G H N V Y N S T
 - S T T * S R V I F * I V G T M C T T P Q
 11881 - AATGATATTCTCTTGCAAAAGACACAACATGAAGCTTCGAGAAAGATGGTTCTCTTTG - 11940
 - N D I D L L A K D T T E A F E K M V S L L
 - M I F F L Q K T Q L K L L S R R W F L F C
 - * Y S S C S K R H N * S F R E D G F S F V
 11941 - TCTGTGTTGGTATCCATCGAGGTCGCTAGACATATAAGGTTCTGAGGAATATGTC - 12000
 - S V L L S M Q G A V D I N R L C E E M L
 - L F C Y P C R V L * T L I G C A R K C S
 - C F A I H A G G C C R H * * U V R G N A R
 12001 - GATAACCGTCTACTCTTCAGGCTATTCGCTAGAGATTGTTCTTACATATAIGCC - 12060
 - D N R A T L Q A I A S E F S S L P S Y A
 - I T V L L F R L L L Q N L V L V L R H M P
 - * P C Y S S G Y C F R I * F F T I I C R
 12061 - GCCTATGCCACTGCCAGGCGCTATAGCAGCGCTGTAGCTAATGGGATTCGAGTC - 12120
 - A Y A T T A Q E A Y E Q A V A N G D S E V
 - L M P L P R R P M S R L * L M V I L K S
 - L C H C G G L * A G C S * W * F * S R
 12121 - GTTCTCAAAGTAAAGAAATCTTGAATGGCTAAATCTGAGTTGACCGTAGCT - 12180
 - V L K K L K K S L N V A K S E F D R D A
 - F S K S * R N L * M W L N L S L T V M L
 - S Q K V K B E I F C G * I * V * F * C C
 12181 - GCGATGCAAGCGAACATTGAAAGATGGCGAGTCAAGCTATGACCCAATGACAAACAG - 12240
 - A M Q R K L E K M A D Q A M T Q M Y K Q
 - P C N A S W K R W Q V I R L * P K C T N R
 - H A T Q V G K D G R S G Y D P N V Q T G
 12241 - HCAAGATCTGAGGACAAGAGGGCANAAGTAACACTGCTATGCCAACATGCTCTTCACT - 12300
 - A R S E D K R A V T S A M Q T M L F T
 - Q D L R T R G Q K * L V L C K Q C S S L
 - K I * G Q E G K S N * C Y A N N A L H Y
 12301 - ATGCTTAGGAGACTTGATATGATGCACTAACACAAATTCAACATGCGCTGATGGT - 12360
 - M L R K L D N D A L N N I I N N A R D G
 - C L G S L I M M M H L T T L P S T M R V M V
 - A * E A * * C T * G H Y Q Q C A * W L
 12361 - TGTTGTCACATCACATCACATGACTACAGCAGCCAAACTCATGGTTGCT - 12420
 - C V P L N I I P L T T A A K L M V V V P
 - V F H S T S Y H * L Q Q P N S W L L S L
 - C S T Q H H T I D Y S S Q T H G C C P *
 12421 - GATTATGGTACCTACAGAACACTGAGGATANACCTTACATATGCACTGCACTC - 12480
 - D Y G T Y K N T C D G N T F T Y A S A L
 - I M V P T R T L V M V T P L H M H L H S
 - L W Y L Q E H L * W * H L Y I C I C T L
 12481 - TGGAACATCCAGCAAGTGTGATCGGATAGCAAGATTGTCACCTTGTGAAATTAAC - 12540
 - W E I O Q V V D A D S K I V Q L S E I N
 - G K S S K L L M R I A R L F N L V K L T
 - G N P A S C * C G * Q D C S T * * N * H
 12541 - ATGGCAATTACCAAAATTGGCTTGGCCTCTATGTTACAGCTAAGAGCCRACCA - 12600
 - M D N S P N L A W P L I V T A L R A N S
 - W T I H Q I W L G L L L Q L * E P T Q
 - G Q F T K F G L A S Y C Y S S K S Q L S

FIG. 11 Con't

34/106

12601 - GCTGTAACTACAGATAARTGACTGAGTCAGTAGCAGCACTRCGACAGATGCTCTGTGCC - 12660
 - A V K L Q N N E L S P V A L R Q M S C A
 - L L N Y R I M N * V Q * H Y D R G C P V R
 - C * T T E * * T E S S S T T D V L C G
 12661 - GCYGGTACCCACACAAACAGCTGTACTGTGACATACTTGCTACTATRACATTTGC - 12720
 - A G T T O T A C T D D N A L A Y Y N N S
 - L V P H K Q L V L M T M H L P T I T I R
 - W Y H T N S L Y * * Q C T C L L * Q F E
 12721 - AAAGGGAGGTAGGGTTGCTGCGATTACTATCAGRCRCCARAGATCTAATGGCTAGA - 12780
 - K G G C R R F V L A L L S D H Q D L K W A R
 - R E V G L C W H Y Y Q T T K I S N G L D
 - G R * V C A G I T P R P R S Q M G * I
 12781 - TTCCCCTAAAGGTAGGGTACAGGTACAGTCACATTACACAGACTGTGACACCCTTGTAGGTT - 12840
 - F P K S S D G T G T I Y T E L E P P C R F
 - S L R V M V Q V F T Q N W N H L V G L
 - P * E * W Y R Y N L H R T G T T L * V C
 12841 - GTTACAGRCACACAAAAGGGCTTAAGTGAATATTGACTCTCATCARAGGCCTAAC - 12900
 - V T D T P K V P K V K Y L Y F I K G L N
 - L Q T H Q K G L K * N T C T S S S K A * T
 - Y R H T K R A * S E I L V L H Q R L K Q
 12901 - AACCTAAATAGAGGTATGGCTGGCGAGTTGGCTGCTACAGTAGCTGCTTCAGGCTGGA - 12960
 - N L R G M V L G S L A T V R L Q A G
 - T N * I E W N C W A V * L L Q Y V F R L E
 - P K * R Y G A G Q Q F S C Y S T S S G W K
 12961 - ATAGTCACAGAAGTCTGGCANTTCACAGTCGCTTCCTCTGTGCTTTGAGTAGAC - 13020
 - N A T E V P A N S T V L S F C A F A V D
 - M L Q K Y L P I Q L C F P S V L L Q * T
 - C Y R T C Q F N C A F L L C F C S R P
 13021 - CCTGCTAANGCATATAAAGGATACCTAGCAAGTGGAGGACARCCAATACCAACTGTG - 13080
 - P A K A Y K D Y L R S G G Q P I T N C V
 - L L K H I R I T * Q V E D N Q S P T V *
 - C * S I * G L P S K W R T T N H Q L C E
 13081 - AAGATGTGTGTACACACAGTGTACAGGACAGGCAATTACTGTGACACAGAGGTAAAC - 13140
 - K M L C H T G T G Q R A I T V T F E A N
 - R C C V H T L V Q O D R Q L L * H Q K L T
 - D V V Y T H T W Y R T G N Y C N T R S * H
 13141 - ATGGRCRCAAGGTCTTGGGTGGCTCATGTGCTGTATTGTAGATGCCACATTGAC - 13200
 - M D Q Z E S F G G A S C C C L Y C R C H I D
 - W T K S P L V L V L H V V C I V D A T L T
 - G P R V L M S S V L * M F H * P
 13201 - CATCCAAATCTAAAGGTATCTGTGACTGAAAGGTAAAGTACGCTCAAATACCTTACCAT - 13260
 - H P N P K G C D L K G K Y V Q I P T T
 - I Q I L K D S V T * K V S T S K Y L P L
 - S K S * R I L * L E R * V R P N T Y H L
 13261 - TGTCATAGACCCAGTGGTTTACATTAAGAACACAGCTGTACCCCTCGGGATG - 13320
 - C A N D P V G F T L R N T V C T V C G M
 - V L M T Q W V L H L E T Q S V P S A E C
 - C * P S G F Y T * K H S L Y R L R N V
 13321 - TGGAAAGGTATGGCTGTAGTTGTGACCAACTCCGGACACCCCTGTGTCAGTCTGGGGAT - 13380
 - W K G Y G C S C D Q L R E P L M Q S A D
 - G K V M A V V V T N S A N P * C S L R M
 - S R L W L * L * P T P R T L D A V C G C
 13381 - GCATCACGTTTTAACGGGTTGGGGTAAGTGCAGCCGCTTACCCGTGCGGA - 13440
 - A S T F L N G F A V * V Q P V L H R A A
 - H Q R F * T G L R C K C S P S Y T V R H
 - I N V F K R V C G V S A A R L T P C G T

FIG. 11 Con't

35/106

13441 - CAGGCACACTAGTACTGTGCTCTACAGGGCTTGTATATTACAACGAAAAAGTCGTG - 13500
 - Q A L V L M S S T G L L I F T T K K V L
 - R H * Y * C R L Q G F * Y L Q R K K C W
 - G T S T D V V Y R A F D I Y N E K S A G
 13501 - TTTTGCAAAAGTCCCTAAAACATATTGCTGCTGCCCTCCAGGAGAAGGATGAGGAARGCA - 13560
 - V L Q S S * K L I A V A S R R R M R K A
 - F C K V P K N * L L S L P G E G * G R Q
 - F A K F L K T N C C R F Q E K D E E G N
 13561 - ATTTATTAAGCTCTACTTTGAGTTAACAGGCATACTATGCTAACATCCAAACATGAG - 13620
 - I Y * T L T L * L R G I L C L T T N M K
 - F I R L L L C S * E A Y Y V * L P T * R
 - L L D S Y F V V K R H T M S N Y Q H E E
 13621 - AGAGTATTTATAACTTGTAAAGATTGCTCAGGGTTGCTGCATGACTTTTCAGT - 13680
 - R L F I T W L K I V Q R L L S M T F S S
 - D Y L * L G * R L L S S G C C P * L F Q V
 - T I Y N L V K D C P A V A V H D F F K F
 13681 - TTAGAGTAGATGTCGACATGGTACCGCATATATCACGTCAGCGTCAACTAAACACAA - 13740
 - L E * M V T W Y H I Y H V S V * L N T Q
 - * S R W * H G T T Y I T S A S N * I H N
 - R V D G D M V P H I S R Q R L T K Y T M
 13741 - TGGCTGATTAGCTATGCTATGCTACGTCATTTGGTAGGGTAAATGTGATACATTAAG - 13800
 - W L I * S H L Y V I L M R V I V I H * K
 - G * F S L C S T S F * * G * L * Y I K R
 - R D L V Y A L R H F D E G N C D T L K E
 13801 - AAATCTCTCCTACATCTGCTGTAGATGTTATTCATAAAGAGGATTGGTAG - 13860
 - K Y S S H T I A V V M M I I S I R R I G M
 - N T R H I Q L L * * * L F Q * E G L V *
 - I L V T Y N N C C D D Y F N K K D W Y D
 13861 - ACTTCGTAGAGARTCTGACATCTACGCGTATATGCTAATTAGGTAGCGTAC - 13920
 - T S * R I L T S Y A M L T * V S V Y A
 - L R R E S * H L T R I C * L R * A C T P
 - F V E N P D I L R V Y A N L G E R V R Q
 13921 - AATCATTTAAAGACTGTCAACTCTGGATGCTATGGTAGGGCATTTGGCG - 13980
 - N H Y * R L Y N S A M L C V M Q A L * R
 - I I K D C T I L R C Y A * D R H C R R
 - S L L K T V Q F C D A M R D A G I V G V
 13981 - TACTGACATTAAGATAATCAGGATCTTATGGAAACTGGTAGGATTGGTAGCTAC - 14040
 - Y * H * I I R I L M G T G T I S V I S Y
 - T D I R * S G S * W E L V R F R * F R T
 - L T L D N Q D L N G N W Y D F G G D F V Q
 14041 - AAGTAGCCACCGGGCTGGGAGCTCTTATGGATCATATTACTCATTTGCTGATGCCA - 14100
 - K * H Q A A E F L L W I H I T H C P
 - S S T R L R S S Y C G F I L L I A D A H
 - V A P G R C G V P I V D S Y Y S L L M P I
 14101 - TCCCTCACTTGTACTGGGCAATTGGCTGCTGAGTCCCATATGATGTCATGGCAAC - 14160
 - S C L * L G H W L L S P I W N L I S Q N
 - P H F D * G I G C * V P Y G C * S R K T
 - L T L T R A L A A E S H M D A D L A K P
 14161 - CACTTATTAAGTGGGATTGGTGAATATGATTTACGGAGAGGAGCTTGTCTCTG - 14220
 - H L L S G I C * N M I L R K R D F V S S
 - T Y * V G F A E I * F Y G R E T L S L R
 - L I K W D L L K Y D F T E E R L C L F D
 14221 - ACCGTTATTAAATATTGGACCAAGACATACATCCCAATTGTTAATACGTTGGATG - 14280
 - T V I L N I G T R H T I P I V L T V W M
 - P L F * I L G P D I P S Q L Y * L F G *
 - R Y F K Y W D Q T Y H F N C I N C L D D

FIG. 11 Con't

36/106

14281 - ATAGGTGTATCCCTCATGGCAAATTAAATGTTTACTGTGTTCCACCTA - 14340
 - I G V S F I V Q T L M C Y F L L C F H L
 - * V Y P S L C K L * C V I F Y C V S T Y
 - R C I L H C A N E F N V L F S T V F P P T
 14341 - CAAGTTGGACCACTAGTAAGAAAATATTGTAGATGTTGTCCTTTGTTGTC - 14400
 - Q V L D H * E K Y L * M V F L L F D F Q
 - K F W T T S K K N I C R W C S F C C F N
 - S F G P L V R K I F V D G V P F V V S T
 14401 - CTGGATACCATTTCTGTGACTTAGGAGTCGATACATAATCAGGATGAAACTTACATAGCT - 14460
 - L D T I F F V S * E S Y I I R M * T Y I A
 - W I P P F S * V R S R T * S G C K L T * L
 - G Y F R E L G V U N H Q D V N L H S S
 14461 - CGCGTCGTCAGTTCAAGGAACCTTGTAGTATGCTGCTGATCCACGATTCGACGCTT - 14520
 - R V S V S R N F * C M L L I Q L C M O L
 - A S Q F Q G T F S V S V C C * S S Y A C S F
 - R L S F K E L L V Y A A D P F A M H A R S
 14521 - CTGGCAATTATTTGCTAGATAACCCACTACATGGCTTCACTAGCTGCCTACAAACCA - 14580
 - L R A Y C * I N A L H A F Q * L H * Q T
 - W Q F I A R * T H Y M L F S S C T N K Q
 - G N L L L D K R T T C F S V A A L T N N
 14581 - ATGTTGCTTCAACATGCAACCCGGTAATTATTAAGACTTTATGACTTTGCTG - 14640
 - M L L F K L S N P V I L I K T F M T L L
 - C C F S N C Q T R * F * * R L L * L C C
 - V A F Q T V K P G N F N K O F Y D D F A V
 14641 - TGTCTAAGGTTCTTAAAGGAGGAGTCTGTTGAACTAAACACTTCTTCTGTC - 14700
 - C L K V S L R K E V L L N * N T S S L L
 - V * R F L * G R K F C * T K T L L L C S
 - S K G F F K E G S S V E L K H F F F A Q
 14701 - AGGATGCCACGCTCATACGTGATTGACTATATCGTTATAATCTCCAACAAATG - 14760
 - R M A T L L S V I M T I I V I I C O Q C
 - G W Q R C Y Q * L * L L S L * S A N N V
 - D G N A A I S D Y D Y Y R Y N L F T M C
 14761 - GTGATPATCAGACAACCTCTATCGTAGTGAAGTGTGATAAFACTTTGATTGTCAG - 14820
 - V I S D N S * L K L L I N T L I V T
 - * V Q T T P I R S * S C * * I L * L L R
 - D I R Q L L F V V E V W D K Y F D C Y D
 14821 - ATGGTGGCTGATATGCAACCARGTARTCGTAPACATCTGGATAAACAGCTGGTT - 14880
 - M V A V L L M P T K * S L T I W I N Q L V
 - W W L Y * C Q P S N R * Q S G * I S W F
 - G G C I N A N Q V I V V N N L D K S A G F
 14881 - TCCCATTTATAATAGGGTAAGCTGAGACTTTATATGACTCATGAGTATGAGGATC - 14940
 - S H L I N G V R L D F M T Q * V M R I
 - P I * * M G * G * T L L * L N E L * G S
 - P F N K W G K A R L Y Y D S M S Y E D Q
 14941 - AAGATGCACITTCGGTAACTAACGGTATAGTCATCCCACATATACTCAATGATC - 15000
 - K M H F S R I L L S V M S S L L * L K * I
 - R C T F R V Y * A * C H P Y Y N S N E S
 - D A L F A Y T K R N V I P T I T Q M N L
 15001 - TTAAGTATGCCATTAGTCAGAAATAGAGCTCGACCGVAGCTGCTGTTCTATCCTGTA - 15060
 - L S M P L V Q R I E L A P * L V S L S V
 - * V C H * C K E * S S H R S W C L Y L *
 - K Y A I S A K N R A R T V A G V S I C S
 15061 - GTACTATGACAAATAGACAGTTCTACGAAATATTGAGTCATAGCCGCACTAGAG - 15120
 - V L * Q I D S F I R N Y * S Q * P P L E
 - Y Y D K * T V S S E I I E V N S R H * R
 - T M T N R Q F H Q K L L K S I A A T R G

FIG. 11 Con't

37/106

15121 - GAGCTACTGGGTAATTGAAACAAAGCAAGTTTACGGTGGCTGGCATAATAAGTTAAAAA - 15180
 - E L L W * L E Q A S F T V A G I I C * K
 - S Y C G N W N N K Q V L R W L A * Y V K N
 - A T V V I G T S K P Y G G W H N M L K T
 15181 - CTGTTACAGTGTAGAAACTCCACACCTTATGGGTGGGATTATTCACAAATGAC - 15240
 - L F T V T V M * K L H T L W V G I I Q N V T
 - C L Q * C R N S T F Y G L G L S K M * Q
 - V Y S D V E T P H L M G W D Y P K C D R
 15241 - GAGCCATGCCATCATGCTTAGGATAANGGCCCTCTGTGTTCTGCTGCAAACATACRA - 15300
 - E P C L T C L G * W P L L F L L A N I T
 - S H A * H A * D N G L S C S C S Q T * H
 - A M P N R R I M A S L V L A R K H N T
 15301 - CTTGCTGTAACTTACACACCGTTCTACAGGTAGCTAACGAGTGTGCGCAAAGTATTAA - 15360
 - L A T V Y H T V S T G * L T S V R K Y *
 - L L * L I T P F L Q V S * R V C A S I K
 - C C M L S H R F Y R L A N E C A R Q V L S
 15361 - GTGAGATGCGICATGCTGGCGCTCACATAGTTAACAGGTGGAACTACATCGCGTG - 15420
 - V R W S C V A A H Y M L N Q V E H H P V
 - * D G H V W R L T I C * T R W N X I R *
 - E M V M C G G S L Y V K P G G T S S G D
 15421 - ATGCCTACACTGCTTACATCTATAGTGTCTTAACTTGTCAAGCTGTTACAGGCCATG - 15480
 - M L Q L M L I V S L T F V K L L Q P M
 - C Y N C L C * * C L * H L S S C Y S Q C
 - R T T A Y A N S V F N I C Q A V T A N V
 15481 - TAAATGCACTTCTTCACAATGCTGGTAATAAGATAGCTGACAAGTATGTCGCCTAC - 15540
 - * M H F F P Q L M V I R * L T S M S A I Y
 - K C T S F N * W * * D S * Q V C P Q S T
 - N A L L S T D G N K I A D K Y V R N L Q
 15541 - AACACAGGCTCTATGAGGCTCTATAGAAATAGGGATGTTGATCATGAATTCTGGATG - 15600
 - N T G S M S V S I E I G M L I M N S W M
 - T Q A R L * V S L * K * G C * S * I R G *
 - H R L Y E C L Y R N R D D V D H E F V D E
 15601 - AGTCTTACGCTTACCTGGGTAACATTTCTCCATGATGATTCTTCATGATGTCGGTIG - 15660
 - S F T L T C V N I S P * F F L M M P L
 - V V L R L P A * T F L R H D D S F * * C R C
 - F Y A Y L R K H F S M M I L S D D A V V
 15661 - TGTCGCTAACACTAACATGCGGCTAACGGTTAGTAGCTAGCATTAAGAACATTAAAGG - 15720
 - C A I T V T M R L K V * L A L R T L R
 - V L * Q * L C G S R F G S * H * E L * G
 - C Y N * A Q G L V A S I K N F K A
 15721 - CAGTTCTTATTATCAAATAATGTTCTATGCTGAGGCAAATGTTGGACTGAGACTG - 15780
 - Q F F I I K I M C S C L R Q N V G L R L
 - S S L L S K * C V H V * G K M L D * D *
 - V L Y Y Q N N V F M S E A K C W T E T D
 15781 - ACCTTACAAAGGACCTACGAAATTTCCTCACACGATAAACATGCTAGTTAACAGGGAG - 15840
 - T L L K D L T N F A H S I Q C * L N K E
 - P Y * R T S R I L L T A Y N A S * T R R
 - L T K G P R E F C S Q H T M L V K Q G D
 15841 - ATGATTACGTTACCTGGCTTACCCAGATCCTACAGAAATATTAGGGCCAGGCTGTTTG - 15900
 - M I T C T C L T Q I H Q E Y * A Q A V L
 - * L R V P A L P R S I K N I R R R L F C
 - D Y V Y L E F Y P D P S R I L G A G C F V
 15901 - TOGATGATATTGTCAAAAGAGTGGTACACTTATGATTGAAAGGTTGGCTGACTGGCTA - 15960
 - S M I L S K Q H V H L * L K G S C H N L
 - R * Y C Q N R W Y T Y D * K V R V T G Y
 - D D I V K T D G T L M I E R F V S L A I

FIG. 11 Con't

38/106

15961 - TTGATGCTTACCCACTTACAAACATCCTAATCAGGAGTATGCTGATGTCCTTCACTGT - 16020
 - L M L T H L O N I L I R S M L M S F T C
 - * C L P T Y K T S * S G V C * C L S L V
 - D A Y P L T K H P N Q E Y A D V F H L Y
 16021 - ATTTACATACATTAGAAGATTCATGATGAGCTTACTGGCCACATGGTGGACATGTATT - 16080
 - I Y N T L E S Y M M S L L A T C W T C I
 - F T I H * K V T * * A Y W P H V G H V F
 - L Q Y I R K L H D E L T G H M L D M Y S
 16081 - CGGAACTGTAACATGATAACACCTCACGGTACTGGGAACTGGAGTTATGAGGCTA - 16140
 - P * C * L M I T P H P G T G N L S F M R L
 - R N A N * * * H L T V L G T * V L * G Y
 - V M L T N D N T S R Y W E P E F Y E A M
 16141 - TGACACACCAACATACAGTCTTGAGGCTGTAGGTGCTGTGTATTGCRATTCAAGA - 16200
 - C T H H I Q S C R L * V L V Y C A I H R
 - V H T T Y S L A G C R C L C I V Q F T D
 - Y T P H T V L V Q A V G A C V L C N S Q T
 16201 - CTTCACCTCGTGGCTGCTGTAGGAGAACCTTCTATGCTGCAAGTGGCTGTATG - 16260
 - L H F V V A V P V L G D H S Y V A S A A M
 - F T S L R C L C L Y * E T I P M L Q V L L *
 - S L R C G A C I R R P F L C C K C C Y D
 16261 - ACCATGTCATTCAACATCACACACACATTAGTGTGCTGTAACTCCATGTTGCAATG - 16320
 - T M S M F Q H H T N * C C L L I P M F A M
 - P C H F N I T Q I S V V C * S L C L Q C
 - R V I S T S H K L V L S V N P Y V C N A
 16321 - CCCAGGTGTGATGACTGATGACACAACACTGTATGAGGATATGAGCTTATT - 16380
 - P Q V V M S L M * H N C I * E V * A I I
 - P R L * C H * C D T T V S R R Y E L L L
 - P G C D V T D V T Q L Y L G G M S Y Y C
 16381 - GCAAGTCRCAATAAGCCCTCCATTAGTGTGCTGTAACTGTCAATGGTCAAGGTTTGGTT - 16440
 - A S H I S L P L V F H Y V L M V R F L V
 - Q V T * A S H * F S I M C * W S G F W F
 - K S H K P P I S F P L C A N G Q V F G L
 16441 - TATACAAACACATGCTGGCAGTCACTGACTGACTGACTGACTGACTGACTGCAACAT - 16500
 - Y T K T H V * A V T M S L S M R * Q H
 - I Q K H M C R Q * Q C H * L S Q C D S N M
 - Y K N T C V G S D N V T D F N A I A T C
 16501 - GTGATGGACTAATGCTGGCATACATGCAACACTGTACTGAGAGACTCAAGC - 16560
 - V I G L M L A I T Y L P T L V L R D S S
 - * L D * C W R L H T C Q H L Y * E T Q A
 - D W T N A G D Y I L A N T C T E R L K L
 16561 - TTTGCGAGCAAAAGCTCAAGCCACTGAGGAAACATTAACTGCTGTCAATGGTATG - 16620
 - F S Q Q K R S K P L R K H L S C H M V L
 - F R S R N A Q S H * G N I * A V I W Y C
 - F A R E T L K A T E E T F K L S Y G I A
 16621 - CCACITGACSCGAAGTACTCTGACAGAACTGCACTTCTATGGAGGTTGGAAAC - 16680
 - P L Y A K Y S L T E N C I F H G R L E N
 - H C T R S T L * Q R I A S F M G G W K T
 - T V R E V L S D R E L H L S W E V G K P
 16681 - CTAGACCAACATTGAGACAGAACTATGCTTTACTGGTTACCGTGTAACTAAAGTAGTA - 16740
 - L D H H * T E T M S L L V T V * L K I V
 - * T T I E Q K L C L Y W N L P C N * K * *
 - R P P L N R N Y V F T G Y R V T K N S K
 16741 - AAGTACAGATTGGAGACTACACCTTGAAGAAAGGTGACTATGGTGTGCTGTGTAC - 16800
 - K Y R L E S T P L K K V T M V M L L C T
 - S T D W R V H L * K R * L W * C C C V Q
 - V Q I G E Y T F E K G D Y G D A V V Y R

FIG. 11 Con't

39/106

16801 - GAGGTACTACGACATACAAGTGAATTGGTGTTACTTGTGTGACATCTCACRGTG - 16860
 - E V L R H T S * M L V I T L C * H L T L
 - R Y Y D I Q V E C W * L L C V D I S H C
 - G T T T Y K L N V G D Y F V L T S H T V
 16861 - TAATGCCACTTAGTCACCACTACTGTGCCAACAGAACACTATGTGAGAATTACTGGCT - 16920
 - * C H L V H L L * C H K S T M * E L L A
 - N A T * C T Y S S A T R A L C E N Y W L
 - M P L S A P T L V P Q B E H Y V R I T G L
 16921 - TGTAACCAACACTCACATCTCAGATGAGTTCTAGCATGTTGCAAATTATCAGG - 16980
 - C T Q H S T S Q M S F L A M L Q I I K R
 - V P N T Q H L R * V F * Q C C K L S K G
 - Y P T L N I S P S N V A N Y Q K V
 16981 - TCGGCATGCCAAAGTCTCACATCCAGTCAGTGTGTTCTAGCATGTTGCAAATTATCAGG - 17040
 - S A C K S T L H S K D H L V L R V I L
 - R H A K V L Y T P R T T W Y W * E S F C
 - G M Q K Y S T L Q G G P P G T G K S H F A
 17041 - CCATCGCCTGCTCTATACCACATCTGCTGAGTGTATACGGCATGCTTCATG - 17100
 - P S D L L S I T H L L A * C I R H A L M
 - H R T C S L L P I C H S V Y G M L S C
 - I G L A L Y Y P S A R I V Y T A C S H A
 17101 - CAGCTGTTGATGCCATGTGAAAAGCATAAAATATTGCCCATAAGATAAATGCTAGTA - 17160
 - Q L L M F Y V K R H * N I C P * I N V V
 - S C * C P M * K G I K I F A H R * M * *
 - A V D A L C E K A L K Y L P I D K C S R
 17161 - GAATCATCTCGGGCTGCCGGTAGAGCTGTTTGTATAAATTCAGAAGTGAATTCAAC - 17220
 - E S Y L R V R A * S V L I N S K * I Q H
 - N H T C A C A R R R V F * * I Q S E F N T
 - I I P A R A R V E C F D K F K V N S T L
 17221 - TAGAACAGTATGTTCTGCACTGTAAATGCAATGCCAGAACACTGCTGACATGTAG - 17280
 - * N M S F S A L * M H C Q K Q L L T L *
 - R T V C F L H C K C I A R N N N C * H C S
 - E Q Y V F C T V N A L P E T T A D I V V
 17281 - TCTTTGAGAATCTCTATGGCTACTTATATGACTGACTGTTGTCATGCTAGACTC - 17340
 - S L M K S L W L L I M T * V L S M L D F
 - L * * N L Y G Y * L * L E C C Q C * T S
 - F D E I S M A T T N Y D L S V V N A R L R
 17341 - GTGCCAACACTACGCTCATATGGCGATCCTGCTCATTAACAGCCCCGGACATG - 17400
 - V Q N T T S I L T A I L L N Y Q P P A H C
 - C K T L R L Y W R S C S I T S P P H I A
 - A K H Y V Y I G D P A Q L P A P R T L L
 17401 - TGACTAAAGCACAACGAAACAGAACATATTTAAATCGAGTGTGCAAGACTATGAAACAA - 17460
 - * L K A H * N Q N I L I Q C A D L * K Q
 - D * R H T R T R I F * F S V Q T Y E N N
 - T K G T L E P E Y F N N S V C R L M K T I
 17461 - TAGGTCAGACATGTTCTTGAACTTGTGCCGTTGCTCTGCTGAAATTGTGACACTG - 17520
 - * V Q T C S L E L V A V V L D K L L T L
 - R S R H V P W N L S P L S C * N C * H C
 - G P D M N F L G T C R R C P A E I V D T V
 17521 - TGAGTCCTTAGTTATGACATAAACGCTAAACGACACAGGATAAGTCAGCTCATGCT - 17580
 - * V L * F M T I S * K H T R I S Q L N A
 - E C F S L * Q * A K S T Q G * V S S M L
 - S A L V Y D N K L K A H K D K S A Q C F
 17581 - TCAAAATGTTCTACAAAGCTGTTATTACACATGATGTTCTGCAATCAACAGACCTC - 17640
 - S K C S T K V L L R H M M F H L Q S T D L
 - Q N V L Q R C Y Y T * C F I C N Q Q T S
 - K M F Y K G V I T H D V S S A I N R P Q

FIG. 11 Con't

40/106

- 17641 - AAATAGCCGTTGTAAGGAAATTCTTACCGCACTCGCTGGAGAAAAGCTGTTTTA - 17700
 - K * A L * E N F L H A I L L G E K L F L
 - N R R C K R I S Y T Q S C L E K S C F Y
 - I G V V R E F L T R N P A W R K A V F I
 17701 - TCTCACCTTATAATCAGAACGCCGTAGCTTCMAAAATCTAGGATTCGCCTACGCAGA - 17760
 - S H L I I H R T L * L Q K S * D C L R R
 - L T L * F T E R C N S F K N L R I A Y A D
 - S P Y N S Q N A V A S K I L G L P T Q T
 17761 - CTGTTGATTCATCACAGGGTCTGTRATATGACTATGTCATATTCACACAARACTTGAAA - 17820
 - L L I H E R V F V L N M T M S Y S H K L L K
 - C * F I T G F * I * L C H I H T N Y * N
 - V D S S Q G S E Y D V V I F T Q T T E T
 17821 - CAGCACACTCTGTAATGTCACCGCTTCATGGCTATCACAGGGCMAAAATTGGCA - 17880
 - Q H T L V M S T A S M W L S Q G Q K L A
 - S T L L * C Q P L Q C G Y H K G K N W H
 - A H S C N V N V R F N V A I T R A K I G I
 17881 - TTTTGTGATATAIGTCGAGCATCTTATGACAATCTGAACTTCATTCAGCTAGAAA - 17940
 - F C A * C L I E E I F M T N C N L Q V * K
 - F V H N V * * R S L * Q T A I Y K S R N
 - L C I M S D R D L Y D K L Q F T S L E I
 17941 - TACCACTGCGAAATGTCGCTCATTCAGCAGAAATGTACTGGACTTTTARGGACT - 18000
 - Y H V A M N L H Y K Q K M * L D F L R T
 - T T S Q C G Y I T S R K C N W T F * G L
 - P R R N V A T L Q A E N V T G L F K D C
 18001 - GTAGTAAAGCATTCATTGCTGCTTCATCACAGGCCACTCACACCTCGCGGTGATA - 18060
 - V V R S L L V F I L H R H L H T S A L I
 - * * D H Y W S S S Y T G T Y T P Q R * Y
 - S K I I T G L H P T Q A P T H L S V D I
 18061 - TAAATTCAGACTGAGGATATGTGTTGACATACCGGCATACCAAGGGACATGACCT - 18120
 - * N S R L K D Y V L T Y Q A X Q R T * P
 - K I Q D * R I M C * H T R H T K G H D L
 - K F K T E G L C V D I P G I P K D M T Y
 18121 - ACCGAGACTCATCTATGTTGGCTCAAATGAATTACAGTCATGTTACCCCTA - 18180
 - T V D S S L * W V S K * I T K T S M V T L
 - P * T H L Y D G F Q N E L P S Q W L P *
 - R R L I S M M G F K M N Y Q V N G Y P N
 18181 - ATATGTTATCACCGCGAAGAAGCTTCTGTCAGCTGGATTCGCTTGTGATG - 18240
 - I C L L S P F A K K L F V T F V R G L A L M
 - Y V Y H P R R S Y S S R S C V D W L * C
 - M F I T A R H V R A W I G F D V
 18241 - TAGAGGGCTGTCATGCAACTAGAGATGCTGTTGGTACTAACCTTCCAGCTAGGAT - 18300
 - * R A V M Q L E M L W V L T Y L S S * D
 - R G L S C N * R C C G Y * P T S P A R I
 - E G C H A T R D A V G T N L P L Q L G F
 18301 - TTCTACAGGTGTAACCTAGTAGCTGTCAGGACTGCTGTTATGTCACACTGAAAATACCA - 18360
 - F L Q V L T * * L Y R L V M L T L K I T
 - F Y R C * L S S C T D W L C * H * K * H
 - S T G V N L V A V P T G Y V D T E N N T
 18361 - CAGAAATTACCAAGAGTAAATGCAAAACCTCCACCAAGTGACCAAGTTAACATCTTAC - 18420
 - Q N S P E L M Q N L H Q V T S L N I L Y
 - R I H Q S * C K T S T R * P V * T S Y T
 - E F T R V N A K P F P P G D Q F K H L I P
 18421 - CACTCATGATAAAGCTGGCTGGAAATGACTCGTATTAGATAGTACAAATGCTCA - 18480
 - H S C I K A C P G M * C V L R * Y K C S
 - T H V * R L A L E C S A Y * D S T N A Q
 - L M Y K G L P W N V V R I K I V Q M L S

FIG. 11 Con't

41/106

18481 - GTGATACACTGAAAGGTCTCGACAGAGTCGTGTTGGCGCATGGCTTG - 18540
 - V I H * K D C Q T E S C S S F G R M A L
 - Y T E R I V R Q S R V R P L G A W L *
 - D T L K G L S D R V V F V L W A H G F E
 18541 - AGCTTACATCAAATGAAGTACTTGTCAAGATTGGCCCTGAAGAGAACGTTGTCGIGTG - 18600
 - S L H Q * S T L S R L D L K E R V V C V
 - A Y I N E V L C Q D W T * K N V L S V *
 - L T S M K Y F V K I G P E R T C C L C D
 18601 - ACAACGTCAGAATTGCTTTACTTCATCAGATACTTATGCCGCTGGAAATCATTG - 18660
 - T N V Q L D A F L L H Q I M P A G I I L
 - Q T C N C L L F Y F I R Y L C L L E S F C
 - K R A T C F T S S D T Y A C W N H S V
 18661 - TGGGTTTGACTATGCTATAACCCATTATGATGATGTCAGCAGTGGGCTTACGG - 18720
 - W V L T M M S I T H L * L M F S S G A L R
 - G F D Y V Y N P F M I D V Q Q W G F T G
 - G F D Y V Y N P F M I D V Q Q W G F T G
 18721 - GTAACCTTCAGAGTCRCCRACACATGGCAGGTACATGGAAATGACACATGGCTA - 18780
 - V T F R V T M T N I A R Y M E M H M W L
 - P S E * P * P T L P G T W K C T C G *
 - N L Q S N H D Q H C Q V H G N A H V A S
 18781 - GTTGTGATGCTATCATGACTAGTGTGTTAGCTCCATGAGTCGTTTAACCGGGTGT - 18840
 - V V M L S * L D V * Q S M S A L L S A L
 - L * C Y H D * M F S S P * V L C * A R *
 - C D A I M T R C L A V H E C F V K R V D
 18841 - ATTGGCTCTGGATAATCCTATTAGGAGATGACAATGAGGTAAATCTGCTTCAGAGA - 18900
 - I G L L N T T L L * E M N * G L I L L A E
 - L V C * I P Y Y R R * T E G * F C L Q K
 - W S V E Y P I D E G D E L R V N S A C R K
 18901 - AACATCACACATGGTGTGAGTCATGCTGATGATAAGTTCCAGTTCTCATG - 18960
 - K Y N T W L * S L H C L L I S F O F F M
 - S T T H G C E V C I A C * * V S S S S *
 - V Q H M V V K S A L L A D K F P V L H D
 18961 - ACATTTGAAATCCACAGGCATTCAGATGTGTGCTCAGGTGAGTAGAATGGAAAGTTCT - 19020
 - T L E I Q R L S S V C L R L K * N G S S
 - H W K S K G Y Q V C A S G * S R M E V L
 - I G N P K A I K C V P Q A E V E W K F Y
 19021 - ACGATGCTCAGGCATCTAGTCAGACAAAGTACAAATAGAGGAACTCTCTTCTTCTAAG - 19080
 - T M L S H V V T K L T K * R N S S I L M
 - R C S A M * * Q S L Q N R G T L L F L C
 - D A Q P C S D K V K I E E L F Y S Y A
 19081 - CTACACATCACATGATAATTCTCATGATGGTTGTTGGAAATTSACGGTGTAC - 19140
 - L H I T I N S L M V F V C F G I V T L I
 - Y T S R * I H * W C L F V L E L * R * S
 - T H H D K F T D G V C L F W N C N V D R
 19141 - GTTACCCAGCCATGCAATGTTGTTAGCTGACACAGAGTCATGCAACTTGACT - 19200
 - V T Q P M O L C V G L T Q E S C Q T * T
 - L P S Q C N C V * V * H K S L V K L B L
 - Y P A N A I V C R F D T R V L S N L N L
 19201 - TACCCAGCTGATGTTGTTGTTGATGAAAGCATGCACTCCACCTCCAGCTT - 19260
 - Y Q A V M V V V C N * I S M H S T L Q L
 - T R L * W W * F V C E * A C I P H S S F
 - P G C D G G S L Y V N K H A F H T P R F
 19261 - TCGATAAAAGTCATTAATTAAGCAATTGCCCTTCTTACTATCTGATAGTC - 19320
 - S I K V H L L I * S N C L S F T I L I V
 - R * K C I Y * F K A I A F L L L F * * S
 - D K S A F T N L K Q L P F F Y Y S D S P

FIG. 11 Con't

42/106

- 19321 - CTTGTGAGTCATGGCARACRAAGTAGTGTGGATATTGATTATGTTCCACTAATCTG - 19380
 - L V S L M A N K * C R I L I M F H S N L
 - L * V S S W O T T S S V G Y * L C S T Q I C
 - C E S H G K Q V V S D I D Y V P L K S A
 19381 - CTAGGTGATTACACGATGCAATTAGGTGGTGCCTTGCAGRCACCATGCAAATGCT - 19440
 - L R V L H D A I * V V V L F A D T M Q M S
 - Y V Y T T M Q F R W C C L Q T P C K * V
 - T C I T R C N L G G G A V C R H A N E Y
 19441 - ACCGACAGTACTTGATGCTATAATATGATGATTTCTGCCTGATTTACCTATGGAT - 19500
 - T D S T W M H I I * F L L D L A Y G F
 - P T V L G C I * Y D D F C W I * P M D L
 - R Q Y L L D A Y N M M I S A G F S L W I Y
 19501 - ACAAACATTTGATACTTATACTCTGGAATACATTACCGAGTTACAGAGTTAGAAA - 19560
 - T N N L I L I T C G I H L P G Y R V * K
 - Q T I * Y L * P V E Y Q V T E F R K
 - K Q F D T Y N L W N T F T R L Q S L E N
 19561 - ATGTCGCTTATATGTTGTAATAAGGACACTTGTGATGGACAGCCGCCGAACCTCG - 19620
 - M W L D M L L I K D T L M D T P A K H L
 - C G L * C C * R T L L * W T T R R R S T C
 - V A Y N V V N K G H F D D G H A G E A P V
 19621 - TTTCATCATTATAATGCTGTTACAAAGGTAGATGGTATTGATGTTGGAGATCTTG - 19680
 - F P S L I M D F T Q R * M V L M W R S L
 - F H H * C C L H K G R W Y * C G D L *
 - S I I N N A V Y T K V D G I D V E I F E
 19681 - AAAATAAGCACACACTTCTGTATAATGCTTGGCTAACGCTAACATTA - 19740
 - K I R Q H F L L M L H D S F G L S V T L
 - K * D N T S C * C C I * A L G * R * H *
 - N K T T L P V N V A F E L W A K R N I K
 19741 - AACCACTGGCAGAGATTAAGACTCTATAATTGGGTGTTGATATCGCTGCTAATCTG - 19800
 - N Q C Q R L R Y S I I W V L I S D L I L
 - T S A R D * D T Q * F G C * Y R C * Y C
 - P V F E I K I L N N L G V D I A A N T V
 19801 - TAATCIGGAGCTACAAAAGAGAACGGCCCAGCACATGTATCTACAAATAGGTGCTGGACAA - 19860
 - S G T K E K P Q H M Y L Q * V S A Q
 - N L G L Q K R S P S T C I Y N R C L H N
 - I W D Y K R E A P A H V S T I G V C T M
 19861 - TGACTGACATGGCAAGAACACTACTGAGAGTGCCTGTTCTACTTACTGTCTGTTG - 19920
 - * L T L P P N N L L R L V L H L D S C L
 - D * H C Q E T Y * E C L F F T Y C L V *
 - T D I A T P T E S A C S S L T V L F D
 19921 - ATGGTAGCTGGAGGACAGGTAGACCTTTAGAAACGCCGTATGGTTTTAAATA - 19980
 - M V E W K D R * T F L E T P V M V F * *
 - W * S G R T G R P F * K R P * N C F N N
 - G R V E G Q V D L F R N A R N G V L I T
 19981 - CAGAAGGTTCACTGAAAGGCTTACACCTTCAAGGGACAGCACAGCTAGCGCTATG - 20040
 - Q K V Q S K V * H L Q R D Q H K L A S M
 - R R F S Q R S N T F K G T S T S * R Q W
 - E G S V K G L T P S K G F A Q A S V N G
 20041 - GAGTCACATTAATGGAGAGATCACTGAAACACAGTTACTACTTAAAGRAAGTAGACG - 20100
 - E S H * L E N Q * K H S L T T L R K * T
 - S H I N W R I S K N * L L * E S R R
 - V T L I G E S V K T Q F N Y F K K V D G
 20101 - GCATTTACACAGTTGCTGAAACCTACTTACTCTAGAGCAGAGACTTAGAGGATTAA - 20160
 - A L F N S C L K P T L L R A E T * R I L
 - H Y S T V A * N L L Y S E Q R L R G F *
 - I I Q Q L P E T Y F T Q S R D L E D F K

FIG. 11 Con't

43/106

20161 - AGCCCAGATCRAAATGGAAACTGACTTCTGAGCTCGTATGGATGAAATTCAACAGC - 20220
 - S P D H K W K L T F S S S L W M N S Y S
 - A Q I T N G N * L S R A R A Y G * I H T A
 - P R S Q M E T D F E L A M D E F I O R
 20221 - PRATAAGCTCGAGGGCTATGCCCTTCGAACACATCGTTATGGAGATTTCAGTCATGGC - 20280
 - D I S S R A M P S N T S F M E I S V M D
 - I * A R G L C L R T H R L W R F Q S S W T
 - Y K L E G Y I R F E H I V Y G D F S H G Q
 20281 - AACATGGGCTCTICATTAATGATAGGCTTGCACAGGCTCACAGRITCACCACTTA - 20340
 - N L A V F I * * * A * P S A H K I H H L
 - T W R S S F N D R L S Q A L T R F T T *
 - L G G L H M I G L K A R R S Q D S P L K
 20341 - AATAGAGGGTTATCCCTATGGACAGCACAGTGGAAAAATTACTCTCATAACAGATGCC - 20400
 - N * R I L S L W T A Q * K I T S * Q M R
 - I R G F Y P Y G Q H S E K L L R H N R C A
 - L E D F I P M D S T V K N Y F I T D R O
 20401 - AACAGGGTTCATCAAATGTGTTGTTGTTGATCTTACTTGAGACTTGTGG - 20460
 - K Q V H Q N V C V L * L I F Y L M T L S
 - N R F I K M C V C F C D * S F T * * L C R
 - T G S S K C V C S V I D L L D D F V E
 20461 - AGATAATAAGTCACAAGATTGTCAGTGATTCTCAAAAGGGTCAAGGGTACAAATTGACT - 20520
 - R * * S H K I C Q * F Q K W S R L Q L T
 - D N K V T R F V S O F K S G Q G Y N * L
 - I I K S Q D L S V I S K V V K U T I D Y
 20521 - ATGCTGAAATTCTTCATGCTTGTGGTGTAAAGGATGGACATGTTGAAACCTTACCCAA - 20580
 - M L K F H S C F G V R M D M L K P S T Q
 - C * N F I H A L V * G W T C * N L L P K
 - A E S F M L W C K D G H V E T F Y P K
 20581 - AACATCAAGCAAGTCAGCTGCAACCAGGTGGTGCATGCCACTGTGCAAGATGC - 20640
 - N Y K Q V K R G N Q V L R C L T C T R C
 - T T S K S S V A T R C C D A * L V Q D A
 - L Q A R S Q A W Q P R G V A M P N L Y K M Q
 20641 - AAAGAATGCTTCTGAAAAGTGTGACCTTCAGAAATTGTTGAAATGCTGTATACCA - 20700
 - K E C F L K S V T F R I M V K M L L Y Q
 - K N A S * K V * P S E L W * K C C Y T K
 - R M L L E K C D L Q N Y G E N A V I P K
 20701 - AAGGATAATGATGATGTCGCAAGTATACTCAACTGTGTCATAACTTAATACACTTA - 20760
 - K E * * * M S Q S I L N C V N T * I H L
 - R N N D E C R K V Y S T V S I L K Y T Y
 - G I M M N V A K Y T Q L C Q Y L N T L T
 20761 - CTTAGCTGCTTCCCTAACACATGAGAGTTATCCTACATTGGTGTGCCTGATAAAGGAG - 20820
 - L * L Y P T T * E L F T L V L A L I K E
 - F S C T L Q H E S Y S L W C W L * * R S
 - L A V P Y N M R V I H F G A G S D K G V
 20821 - TTGCAACGGTCAAGCTGTGCTCACACATGGTGCACACTGGCACACTCTGTGATT - 20880
 - D H Q V Q L C S D N G C Q L A H Y L S I
 - C T R Y S C A Q T M V A N W H T T G C R F
 - A P G T A V L R Q W L P T G T L L V D S
 20881 - CACATCTTAATGACTCTCGCAGCAGATTCTACTTTAATGGAGACTGTGCACACAG - 20940
 - Q I L M T S S P T Q I L L * L E T V Q Q
 - R S * * L R L R R F Y F N W R L C N S
 - D L N D F V S S D A D S T L I G D C A T V
 20941 - TACATACGGCTATAATGGGACCTTATTAATGGCATATGTGACCTAGGACCAAC - 21000
 - Y I R L I N G T L L L A I C M T L G P N
 - T Y G * * M G P Y Y * R Y V * P * D Q T
 - H T A N K W D L I I S D M Y D P R T K H

FIG. 11 Con't

44/106

21001 - ATGTGACAAAAGAGAAATGACTCTAAGAAGGGTTTCACTATCTGTGTGGATTTATAA - 21060
 - M * Q K R M T L K K G F S L I C V D L *
 - C D K R E * L * R R V F H L S V N I Y K
 - V T K E N D S K E G F F T Y L C G F I K
 21061 - AGCAAAAATAGCCCTGGCTGGTTCTATAGCTGAAAGATAACAGAGCATTCCTGGATG - 21120
 - S K N * P N V V W L * L * R * Q S I L G M
 - A K T S P G W F Y S C K D N R A F L E C
 - Q K L A L G G S I A V K I T E H S W N R
 21121 - CTGACCTTACARAAGCTTATGGGCCATTCTCATGGTGGACAGCTTGTACARAGTAA - 21180
 - L T F T S L W A I S H G G Q L L L Q M *
 - * P L Q O A Y G P F L M V W D S F C Y K C K
 - D L Y M G H F S W W T A F V T N V N
 21181 - ATGCATCATCGGAAGCATTTTAATTTGGGGCTAACATCTGGCAAGCGAAGGAC - 21240
 - M H H H R K H F * L G L T I L A S R R N
 - C I I I G S I F N W G * L S W Q A E G T
 - A S S S E A F L I G A N Y L G K P K E Q
 21241 - A A A T T G C T G C A T T A C C A T G C A T G C T A C T A C T T T C T G G A G G A C A C A A R T C T A T C C - 21300
 - K L M A I P C M L T T F S G G T Q I L S
 - N * W L Y H A C * L H F L E E H K S Y P
 - I D G Y T M H A N Y I F W R N T N P I Q
 21301 - AGTTGTCCTCCATTACTCTTGTGACATGACAAATTCTCTTAAATTAAAGAGGAACCTG - 21360
 - S C L P I N S L T * A N F L L N * E E L
 - V V F L F T L * H E Q I S S * I K R N C
 - L S Y S L F D M S K F P L K L R G T A
 21361 - CTGTAATGCTCTTAAAGGNGATCAAATCATGATGATGTTTACTCTCTGGAAAAG - 21420
 - L * C L L R R I K K S M I * F I L F W K K
 - C N V S * G E S S N Q * Y D L F S S G K R
 - V M S L K E N Q I N D M I Y S L L E K G
 21421 - GTAGGCTTATCATGGAGAAACACAGGTGTTCTGATGATATTCTGTTRACA - 21480
 - V G L S S L E K T T E L W F Q V I F L L T
 - * A Y H * R K Q O Q S C G F K * Y S C * Q
 - R L I I R E N N R V V V V S S D I L V N N
 21481 - ACTAAAGCAGCATGTTATTTCTCTATTTCTACTCTCACTAGGGTAGTGACCTTG - 21540
 - T K R T C L F S Y Y F L L L S L V V V T L
 - L N E H V Y F L L I I S Y S H * W * * P *
 - * T N M F I F F L L F L T L T S G S D L D
 21541 - ACCGGTGCACCCCTTGTGATGTTCAAGCTCTAACTCACACATCATCTCATCTA - 21600
 - T G A P L L M M F K L L I T L N I L H L
 - P V H H F * * C S S S * L H S T Y F I Y
 - R C T T F D D V Q A P N Y T Q H T S S M
 21601 - TGAGGGGGTTTACTCTGGATGAAATTCTGATCACAGACTCTTATTTAACATCAGG - 21660
 - * G G F T I L M K F L D Q T L F I * L R
 - E G G L L S * * N F * I R H S L F N S G
 - R G V Y Y P D E I F R F S D T L Y L T Q D
 21661 - ATTATTTCTCCATTTTATCTAATGTTACAGGGTTCTACATATTAACTACATACGGTTG - 21720
 - I Y F F H L M L Q G F I L L I I R L
 - P I S S I L F * C Y R V S Y Y * S Y V V W
 - L F L P F Y S N V T G F H T I N H T F G
 21721 - GCAACCTGTCTACACCTTTAAAGGATGTTATTTCTGCCCAGAGAAATCAGATG - 21780
 - A T L S Y L L R M V F I L L P Q R N Q M
 - Q P C H T F * G W Y L F C C H R E I K C
 - N P V I P F K K D G I Y F A A T E K S N V
 21781 - TTGTCCTGGTTGGTTTGGTTCTACCATGACAAACAGTCACAGTOGGTGTATTATR - 21840
 - L S V V G F L V L P * T T S H S R * L L
 - C P W L G F W F Y H E Q Q V T V G D Y Y
 - V R G W V F G S T M N N K S Q S V I I I

FIG. 11 Con't

45/106

21841 - TTAACAAATTCTACTAATGTTGTATACGRGCATGACTTGTGACAACCTT - 21900
 - L T I L L M L L Y E H V T L N C V T T L
 - * Q F Y * C C Y T S M * L * I V * Q P P F
 - N N S T N V V I R A C N F E L C D N P F
 21901 - TCTTTCTGTTTCTAACCCATGGCTACACAGACACATACATGATATTGGATAATGCT - 21960
 - S L L F L N P W V H R H I L * Y S I M H
 - L C C F * T H G Y T D T Y Y D I R * C I
 - F A V S K P M G T Q T H T M I F D N R F
 21961 - TTAAATGCCACTTCGAGTACATATCTGATGCCCTTCTGCTTGATGTTGAGAAAAGTCAG - 22020
 - L I A L S S T Y L M P F R L M F Q K S Q
 - * L H F R V H I * C L F R A * C F R K V R
 - N C T F E Y I S D A F S L D V S E K S G
 22021 - CTAAATTTAACACTTACCGAGTTGCTTAAATAATAAGATGGTTCTCATGTTT - 22080
 - V I L N T Y E S L C L K I K M G F S M F
 - * F * T L T R V C V * K * R W V S L C L
 - N F K H L R E F V F K N K D G F L Y V Y
 22081 - ATAAGGGCCTTCACCTAACATTAAGATGATGTTCTGATCTACCTCTGGTTTACACTTGA - 22140
 - T R A I N L * M * F V I Y L L V L T L *
 - * G L S T Y R C C S * S T F W F * H F E
 - K G Y Q P I D P V V R D L P S G F N T L K
 22141 - AACCTATTTAACCTTAAAGTGCCTCTGCTTAAACATACAAATTAGAGGCATTCTACAG - 22200
 - N L F L S C L L V L T L Q I L E P F F L Q
 - T Y F * V A S W Y * H Y K F * S H S Y S
 - P I F K L P L G I N I T T N F R A I L T A
 22201 - CCTTTTCACTGCTCAAGACATTGGGCACGTGACTGCGCTTATTGTTGGCTATT - 22260
 - P F H L L K T F G A R Q L Q P I L L L A I
 - L F T C S R H L G H V S C S L F C W L F
 - F S P A Q D I W G T S A A R Y F V G Y L
 22261 - TAAAGGCCAACATCATTTATGCTCAAGTATGATAAAATGTGACAAATCAGATGCTGTG - 22320
 - * S Q L H L C S S M M K M V O S Q M I L
 - K A N Y I Y I A Q V * + K W Y N H R C C *
 - K P T T F M L K Y D E N G T I T D A V D
 22321 - ATGGTTCTCAAATCCTCTGACTCTAACATTGCTCTGTTAGACCTTGAGATTGACA - 22380
 - I V L K I H L L N S N A L L R A L R L T
 - L F S K S T C T * T Q M L C * E L * D * Q
 - C S Q N P L A E L K C S V K S F E I D K
 22381 - AAGGAATTACCGACCTCTAATTCAGGGTGTCTCTCAGGAGATCTGTGAGATTC - 22440
 - K E F I T R P L I S G L F P Q E M L * D S
 - R N L P D L * F Q G C S L R R C C E I P
 - G I * Q T S N F R V V P S G D V V R F P
 22441 - CTAAATTTAACACTTGTCTTGGAGGGTTTAAATGCTACTAAATCCCCTCTG - 22500
 - L I L Q T C V L L E R F L M L L N S L L
 - * Y Y K L V S F W R G F * C Y * I P F C
 - N I T N L C P F F G E V F N A T K F P S V
 22501 - TCTATGCATGGAGRGAARAAARATTCATATGCTTGCTGATTAATCTGCTCTACA - 22560
 - S M H G R E K K P L I V L L I T L C S T
 - L C M G E K K N F * L C C * L L C A L Q
 - Y A W E R K K I S N C V A D Y S V L Y N
 22561 - ACTCACATTTCACCTTAAAGTCTATGCTATGGCTTCTGCCACTRAAGTGAATGATC - 22620
 - T Q H F F Q P L S M A M A F L P L S * M I
 - L N I F F F N L * V L W R F C H * V E * S
 - S T F F S T F K C Y G V S A T K L N D L
 22621 - TTTGCTCTCCATGCTATGCTAGATCTTGTAGTCAAGGGAGATGATGATGACARA - 22680
 - F A S P M S M Q I L L * S R E M M * D K
 - L L L Q C L C R F F C S Q G R * C K T N
 - C F S N V Y A D S F V V K G D D V R Q I

FIG. 11 Con't

46/106

22681 - TAGCGCCAGGACAACTGGTGTATTGCTGATTAATTATTAATTGCCAGATGATTTCA - 22740
 - * R Q D K L V L L L I I I N C Q M I S
 - S A R T N W C Y C * L * L * I A R * F H
 - A P G Q T G V I A D Y N Y K L P D D F M
 22741 - W V V S L L G I L G T L M L L Q L V I I
 - G L C P C L E Y * E H * C Y F N W * L *
 - G C V L A W N T R N I D A T S T G N Y N
 22801 - A T T A T R A T A T A G G T A T C T T A G G C A T G G C A R G C T T A G G C C I T T G A G A G A G A C A T A C T A - 22860
 - I I N I G I L D M A S L G P L R E T Y L
 - L * I * V S * T W Q A * A L * E R R H I *
 - Y K Y R L R H R P F E R D I S N
 22851 - A T G G C T T C T C C C T G A T G G C A A A C C T T G C A C C C C A C C T G C T C T T A A T G T T A T G G C - 22920
 - M C L S P F L M A N L A P H L L L I V I G
 - C A F L P * W Q T L H P T C S * L L L A
 - V P F S P D G K P C T P P P A L N C Y W P
 22921 - C A T A A T G A T I A T G G T T T A C A C C A T C T G G C A T T G G C T A C C A C C T T A C A G G G T G - 22980
 - H * M I M V F T F L L A L A T N L T E L
 - I K H * L W F L R H H Y W H W L P T L L Q S C
 - L N D Y G F Y T T T G I G Y Q P Y R V V
 22981 - T A G T A C T T C T T G A C T T T A A T G C A C C G C C A C G G T T T G G C A C C A A A T T T A T C C A - 23040
 - * Y F F L L N F * M H R P R F V D Q N Y P
 - S T F F * T F K C T G H G L W T K I I H
 - V L S F E L L N A P A T V C G P K L S T
 23041 - C T G A C T T F T T R A G A R C C A G I G T G C A R T T T T A A T T A T G G A C T C A C T G G T A C T G G T G - 23100
 - L T L L R T S V S V S I L I L M D S L L V L V
 - * P Y * E P V C C Q F * F * W T H W Y W C
 - D L K N Q C V N F N G L T G T G V
 23101 - T G T T A C T C C T T C T F C A A A G G A G T T C A C C A T T C A C A R A T T T G G C C G T G T G T T C G - 23160
 - C * L L L Q R D F N H F N N L A V M F L
 - V N S F F K E I S T I S T I W P * C F *
 - L T P E S K R F Q P F Q Q F G R D V S D
 23161 - A T T T C A C T G A T T C G G T C G A G A G T C C T T A A A C T G A A A T T A T G A C A T T C A C C T T G C T - 23220
 - I S L I P F E I L K H L K Y * T F H L A
 - F H * F R S R S * N I * N I R H F T L L
 - F T D S V R D P K T S E I L D I S P C S
 23221 - C T T T T G G G G G T G T A M G T G T A T T A C A C C T G G A C C A A A Y G C T T C A T C T G A A G T T G C T G T C - 23280
 - L L Q D V N C T D V S T A I H A D Q L T P
 - F W G C K C N Y T W N N K C F I * S C C S
 - F G G V S V I T P G T N A S S E V A V L
 23281 - T A T A T C A A G A T G T T A C T G C R C T G M T G T T C A C G A C A T T C A T G C A G A T C A C T C A C R C - 23340
 - Y I K M L T A L M F L Q Q F M Q I N S H
 - I S R C * L H * C F Y S N S C R S T H T
 - Y Q D V N C T D V S T A I H A D Q L T P
 23341 - C A G C T T G C C G C T A T T T C A T C T G G A A C A A T G T A T T C C A G A C T C A A C C A G G C T G T C T T A - 23400
 - Q L G A Y I L L E T M Y S R L K Q A V L
 - S L A H I F Y W K Q C I P D S S R L S Y
 - A W R I Y S T G N N V F Q T Q A G C L I
 23401 - T A G G A G C T G A C A T G T C G A C A C T T C T T A T G A G T G C G A C A T T C C T A T T G G A G C T G C A I T T - 23460
 - * E L S M S T L L M S A T F L L E L A F
 - R S * A C R H F L * V R H S Y W S W H L
 - G A E H V D T S Y E C D I P I G A G I C
 23461 - G T G C T A G T T A C C A T A C A G T T C T T T A T T G A G T G A C T A T G C C A A A A T C T A T T G G C C T T - 23520
 - V L V T I Q F L Y Y V V L A K N L L W L
 - C * L P Y S F F I T * Y * P K I Y C G L
 - A S Y H T V S L L R S T S Q K S I V A Y

FIG. 11 Con't

47/106

23521 - ATATCATGCTTCTAGGTGCTGATACTTCATAATTGCTTAACATTGCTATCTAAACACCAATTGCTATAC - 23580
 - I L C L * V L I V Q L L T L I T P L L Y
 - Y Y V F R C * * F N C C L L * * H H C Y T
 - T M S L G A D S S I A Y S N N T I A I P
 23581 - CTACTRACTTTCAATTAGCATTACTACAGAAGTAATGCCCTGTTCTATGGCTAAAACCT - 23640
 - L L T F Q L A L L Q K * C L F L W L K P
 - Y * F N * H Y Y R S N A C F Y G * N L
 - T N F S I S I S I T T E V M P V S M A K T S
 23641 - CCGTAGATGTAATATGTCATCTCGGGAGATTCTACTGAATGTGCTAATTGCTTC - 23700
 - P * I V I C T S A E S I D L N V L I C F S
 - R R L * Y V H L R R F Y * M C * F A S P
 - V D C N T M Y C I G D S T E C A N L L L Q
 23701 - AATATGGTAGCTTTGCGACACAACTAAATCTGCACTCTCAGGTATTGCTGCTGACAGG - 23760
 - N M V A F A H N * I V H S Q V L L L N R
 - I W * L L H T T K S C T L R Y C C * T G
 - Y G S C F T Q L D N R A L S G I A A E Q D
 23761 - ATCGCACACACAGTGAGATGTTCCCTCAAGTCACACACATGTACAAACCCCCACTTGA - 23820
 - I A T H V V K C S L K S N K C T K P Q L *
 - S Q H T * S V R S S Q T N V Q N P N F E
 - R N T R E V F A Q V K Q M Y K T P T L K
 23801 - AATATTTGGTGGTTTAAATTTCACAAATATTACCTGACCCCTAAAGCCAACTAAGA - 23880
 - N I L L V V L I F H K Y Y L T L * S Q L R
 - I F W W W F * F F T N I T * P S K A N * E
 - Y F G G F N F S Q I L F D P L K E T T K R
 23881 - GGCTTTATTGAGGACTTGTCTTTAAAGGTGACACTCGCTGATGCTGCTCATGA - 23940
 - G L L D L R T C S L I R * H S L M L A S *
 - V F Y * G L A L * * G D T R * C N L H E
 - S F I E D D L L F N K V T L A D A G F M K
 23941 - AGCRATATGGCGAACCTCTAGGTGATATAATGCTAGAGATCTCATTGTGCGCAGAAGT - 24000
 - S N M A N A * V I L M L E I S F V R R S
 - R I W R M P R * Y * C * R S H L C A E V
 - Q Y G E C L G D I N A R D L I C A Q K F
 24001 - TCAATGCTTACAGTGTGGCACCTCTGCCTCACTGATGATGATGCTGCTCCTACTG - 24060
 - S M D L Q C C H L C S L M I * L L P T L
 - Q W T Y S V A T S A H * * Y D C C L H C
 - N G L T V L P P L L T D D M I A R A Y T A
 24061 - CTGCTCTAGTGTAGTGTACTGCACTGCTGGATGRCATTGGCTGGCGCTGCTTC - 24120
 - L L * L V V L P L L D G H L V L L A L L F
 - C S S * W Y C H C W M D I W C W R C S S
 - A L V S G T A T A G W T F G G A G A A L Q
 24121 - AAAFACCTTTGCTAAGCAATGGCATATGGCTCATGGCATGGAGTTACCCAAATG - 24180
 - K Y L L L C K W H I G S M A L E L P K M
 - N T F C Y A N G I * V Q W H W S Y P K C
 - I P F A M Q M A Y R F N G I G V T Q N V
 24181 - TTCTCTATGAGAACCAAACAAATGCCAACCAATTTCACAGGGATTAGTCAAATTG - 24240
 - F S M R T K N K S P T N L T R R L V K F
 - S L * E P K T N R Q P I * Q G D * S N S
 - L Y E N Q K Q I A N Q F N K A I S Q I Q
 24241 - AAGATCACTTACAACACATCACTGCATTGGCAGCTGCAAGNCCTGTTAACCGA - 24300
 - K N H L Q Q H Q L H W A S C K T L L L T R
 - R I T Y N N I N C I G Q A A R R C * P E
 - E S L T T T S T A L G K L Q D V V V N Q N
 24301 - ATGCTCAAGCTTAAACACACTGTTAACACTGATGCTCATTTGGCTGCAATTTCAG - 24360
 - M L K H * T H L L N N L A L I L V Q F Q
 - C S S I K H T C * T T * L * F W C N F K
 - A Q A L N T L V K Q L S S N F G A I S S

FIG. 11 Con't

48/106

24361 - GTGTGCTAATGATATCCTTCGCGACTTGATAAGTCGAGGGGGAGGTACAAATTGACA - 24420
 - V C * M I S F R D L I K S R R R Y K L T
 - C A K * Y P F A T T * * S R G G G T N * Q
 - V L N D I L S R L D K V E A E V Q I D R
 24421 - GGTAAATACAGGCAGACTTCAAAGCCCTCAAAACCTATGTAACACAACTAATCAGGG - 24480
 - G * L Q A D F K E F K P N * H N N * S G
 - V N Y R Q T S K P F S N L C N T T T N Q G
 - L I T G R L Q S L Q T Y V T Q Q L I R A
 24481 - CTGCTGAATCAGGGCTTCTGCTAATCTTGCTGCTACTAAATGTCGAGTGTGTTCTG - 24540
 - L L K S G L L E I L L L K C L S V F L
 - C * N Q G F C * S C C Y * N V * V C S W
 - A E I R A S A N L A T A T K M S E C V L G
 24541 - GRCATCAAAGAGTGGACTTGTGGAAAGGGTACACCTTATGCTCTCCCAAAG - 24600
 - D N O K E L T F V E R A T T L C P S H K
 - T I K K S * L L W K G L P P Y V L P T S
 - Q S K R V D F C G K G Y H L M S F P Q A
 24601 - CAGCCCCGATGGTGTCTCTCATGTCAGCTATGCCCATTCCAGGAGGAGRACT - 24660
 - Q P R M V L L S S Y M S R M C H P R R G T
 - S P A W C C L P T C H V C A I P G E E L
 - A P H G V V F L L H V T Y V F S Q E R N F
 24661 - TCACCAAGCGCAGCRATITGTCATGAAGGCRANGCATACTTCCCTCGTGTAGGTGTT - 24720
 - S F Q R Q Q F V M K A K H T S L V K V F
 - H H S A S N L S * R Q S I L P S * R C F
 - T T A P A I C H E G K R A Y F P R E G V F
 24721 - TTGTTTATGGCATTCTGGTTTATACACAGAGCACTCTTTCTCAAATAA - 24780
 - L C L M A M L L G L L H R G T S F L H K *
 - C V * W H F L V Y Y T E E L L F S T N N
 - V F N G T S W F I T Q R N F F S P Q I I
 24781 - TTACTACAGACATACATTGTCAGGAATTGGATGTCGTTATTGGCATCATAMCA - 24840
 - L L Q T H I L S Q E I V M S L L A S L T
 - Y Y R Q Y I C G L R K L * C R Y W H H * Q
 - T D T F V S G N C D V V I G I I N N
 24841 - ACACAGTTATGATCTCTSCAACCTGAGCTTGTACTCATTCRAAGGAAGAGCTGGACAGT - 24900
 - T Q F M I L C N L S L T H S K K S W T S
 - H S L * S S A T * A * L I Q R R A G Q V
 - T V Y D P L Q P E L D S F K E E L D K Y
 24901 - ACTTCAAAARTCATACATGCCAGATGTGATCTGGCGACATTTCAGGCAATTACCGT - 24960
 - T S K I I H Q M L I L A T F Q A L T L
 - L Q K S Y I T R C * S W R H F R H * R F
 - F K N H T S P D V G D I S G I N A S
 24961 - CTGTCGCTCACATTCAGAAAAGAAATTGACCGCTCATGAGTCGCTAAATTAAATG - 25020
 - L S S T F K K K L T A S M R S L K I * M
 - C R Q H S K R N * P P Q * G R * K F K *
 - V V N I O K E I D R L N E V A K N L N E
 25021 - AACACTCATGACCTTCAGAATTGGGAAATATGACCAATATTAATGCTGGGT - 25080
 - N H S L T F K N N W E N M S N I L D N G L G
 - I T H * P S R I G K I * A I Y * M A L V
 - S L I D L Q E L G K Y E Q Y I K W P W Y
 25081 - ATGTTTGGCTCGGCTTCATGGCTGGACTTATGGCATGTCATGGTTACATCTGGCTT - 25140
 - M F G S A S L L D * L P S S W L Q S C F
 - C L A R R L H C W T N C H R H G Y N L A L
 - V W L G F I A G L I R A I V M V T I L L C
 25141 - GTTCATGACTATGGTGTGGACTTCAGGGCTCATGTCATGGTTACATCTGGCTT - 25200
 - V A * L V V A V A S R V H A L V V L A A
 - L H D * L L Q L P Q G C M L L W F L L Q
 - C M T S C C S C L K G A C S C G S C C K

FIG. 11 Con't

49/106

25201 - AGTTTGATGAGGATGACTCTGAGCCAGTCTCAAGGGTGTCAAATTACATTACACATAAA - 25260
 - S L M R M T L S Q F S R V S N Y I T H K
 - V * * G * L * A S S Q G C Q I T L H I N
 - F D E D D S E P V L K G V K L H Y T * T
 25261 - CGAAGCTTATGGATTGTTATGAGATTTCATCTGGATCAATTACTCGCACAGCAGT - 25320
 - R T Y G F V Y E I F Y S W I N Y C T A S
 - E L M D L F M R F F T L G S I T A Q P V
 - N L W I C L * D F L L D Q L L H S Q *
 25321 - AAAATTGCAATGCTTCTCGCAAGTACTGTCATGCTACAGCAACGATACCCGAC - 25380
 - K N * Q C F S C K Y C S C Y S N D T A T
 - K I D N A S P A S T V Q H A T A T I P L Q
 - K L T M L D L Q V L F M L Q Q R Y R Y K
 25381 - AGGCCTCACCCCTTCGGAGCTTGTATTCGGCTGCTTCTGTGTTTCAGAG - 25440
 - S L T P F R M A C Y W R C I S C C F S E
 - A S L P F G W L V I G V A F L A V F Q S
 - P H S L S D G L L L A L H F L L F F R A
 - CGCTACCRAAAATATTGCGCTCATAAAGATGGCAGTAGCCCTTTAAGGGCTTCC - 25500
 - R Y Q N N N C A Q * K M A A S P L * G L P
 - A T K I I A L N K R W Q L A L Y K G F Q
 - L P K * L R S I K D G S * P F I R A S S
 25501 - GTTCATTTGCAATTACTGCGTATTGGTACCATCTATTACATCTTTGCTGCG - 25560
 - V H L Q F T A R A I C Y H L F T S F A C R
 - F I C N L L L F V T Y I S H L L L V A
 - S F A I Y C C Y L L P S I H I F C L S L
 25561 - TGCAAGGTAAGGAGGCATTTTGACCTCTATGCCCTGATATAATTCTACATGCA - 25620
 - C R * G G A I T F P L C L D I F S T M H
 - A G K E A Q F L Y L Y A L I Y F L Q C I
 - Q V R R R N F C T S M P * Y I F Y N A S
 25621 - CAAAGCTAGTAGATTATTAGAGAATGTTGGCTTGTGGRAACTGCAAACTCAGAAC - 25680
 - Q R M * Y E M L A L L E V Q I Q E P
 - N A C R I I M R C W L C W K C K S K N P
 - T H V E L L * D V G F V G S A N P R T H
 25681 - ATTACTTTATGCCRACACTTGTGGCTGGCACACATARACTATGACTCTGTAT - 25740
 - I T L * C Q L L C D L A H T * L * L L Y
 - L L Y D A N Y F V C W H T R N Y D Y C I
 - Y F M M P T T L F R A G T H I T M T T V Y
 25741 - ACCATATAACAGTGTACAGATCAATTGCTGTTACTGAAGGTGACGGCATTCACACC - 25800
 - T I * Q C H R Y N C R Y * R * R H F N T
 - P Y N S V T D T I V V T E G D G I S T P
 - H I T V S Q I Q L S L L K V T A F Q H Q
 25801 - AAAACTCAAGAGAGACTCAAAATTGGCTTATCTGAGGATAGGCACATCAGGTAA - 25860
 - K T Q R L P N W W L F * G * A L R C *
 - K L K E D Y Q I G G Y S E D R H S G V K
 - N S K K T T K L V V I L R I G T Q V L K
 25861 - AGACTATGCGTGTACATGGCTATTACCGAGTTTACTACCAAGCTTGAAGTCAACCA - 25920
 - R L C R C T W L F H R S L L P A * V Y T
 - D Y V V V H G Y F T P E V Y Y Q L E S T Q
 - T M S L Y M A I S P K F T T S L S L H K
 25921 - ATTACTACAGACACTGGTATGAAATGCTACATTCTCATCTTAAACAGCTGTTAA - 25980
 - N Y Y R H W Y * K C Y I L H L * Q A C *
 - I T D T G I E N A T F F I F N K L V K
 - L L Q T L V L K M L H S S S L T S L L K
 - AGACCCACCGAATGCAAAATCACACATCGACGGCTTCTCAGGGATGCTAATCCAGC - 26040
 - R P T E C A N T H N R R L F R S C * S S
 - D P P N V Q I H T I D G S S G V A N P A
 - T H R M C K Y T Q S T A L Q E L L I Q Q

FIG. 11 Con't

50/106

26041 - AATGGATCCRATTATGATGAGCCGACGACGACTACTAGCGTGCCTTGTAAGCAGA - 26100
 - N G S N L * * A D D D Y * R A F V S T R
 - M D P I Y D E P T T T T S V P L * A Q E
 - W I Q F M M S R R R L D L A C L C K H K K
 26101 - AATGAGTAGCAACTTATGACTCATTCGTTGAGAAACAGGTACGTTAATAGTTAA - 26160
 - K * V R T Y V L I R F G R N R Y V N S *
 - S E Y E L M Y S F V S B E T G T L I V N
 - V S T N L C T H S F R K K Q V R * * L I
 26161 - TAGCGTACTCTTTCTGCTTCGTTGTAATCTGCTAGTCACACTAGCCATCCTTAC - 26220
 - * R T S F S C F R G I S L A S H T S H P Y
 - S V L L F L A F V V F V F L L T L A I L T
 - A Y F F F L L S W C S * S H * P S L L
 26221 - TGCGCTGATTGTCGCTACTGCTGAATATTCTGAGTTAGTTAGTAAACCAAC - 26280
 - C A S I V C V C V L L Q Y C * R E F S K T N
 - A L R L C A Y C C N I V N V S L V K P T
 - R F D C V R T A A I L L T * V * * N Q R
 26281 - GGTTTACGCTTACTGCGCTTAAATACTGAGACTCTCTGAGGGATCTCGATCTCT - 26340
 - G L R L L A C * R S S * R S S * S S
 - V Y V Y V S R V K N L N S S E G V P D L L
 - F T S T R V L K I * T L L K E F L I F W
 26341 - GGTCTRAACGAACTAACTATTATTATTATTCTGTTGGAACITTAACATGCTTATCATG - 26400
 - G L N E L T I I I L F G T L T L L I M
 - V * T N * L L D F C L E L * R C L S W
 - S K R T N Y Y S V W N F N I A Y H G
 26401 - GCAGACACGGTACTATTACCGTGTAGGGAGCTTAAACACTCTGGACACATGAGCTT - 26460
 - A D N G T I T V E E L K Q L L E Q W N L
 - Q T T V T L L P L R S L N N S W N N G T *
 - R Q R Y Y Y R * G A * T T P G T M E P S
 26461 - GTAAATAGGTTCTATTCTAGCCTGGATTATGTTACTAACATTGCTTATCTATCGG - 26520
 - V I G F L L A W A I M L L Q F A Y S N R
 - * * V S Y S * P G L C Y Y N L P I L I G
 - N R F P I P S L D Y V T T I C I D F * S E
 26521 - AACAGGTTTTGATCATATRAAGCTGTTTCTCGCTCTGCGGCACTGACACTT - 26580
 - N R F L Y I I K L V F L W L W P V T L
 - T G F C T * * S L F S S G S C G Q * H L
 - Q V F V E N N K A C F P D L A L V A S N T C
 26581 - GCTTGTGTTGCTGCTGTTGTCACAGAATTATTGGGTGACTGGCGGATTGCGATT - 26640
 - A C F V L L A V V Y R I N N W V T G G I A I
 - L V L C L L L S T E L I G * L A G L R L
 - L F C A C C C L Q N * L G D W R D C D C
 26641 - GCAATGGCTGTTATGAGCTTGATGCTGGCTTAGGACTACTCTGCTTCCCTCAGGCTG - 26700
 - A M A C I V G L M W L S Y F V A S F R L
 - Q W L V L * A * C G L A T S L L P S G C
 - N G L Y C R L D V R * L L R C F L Q A V
 26701 - TTGCTCGTACCCGGCTCATGTGGTCATTCACCCAGAACACATCTCTCATGTG - 26760
 - F A R T R S M W S N P E T N I L L L N V
 - L L V P A Q C G H S T Q K Q T F F S M C
 - C S Y P L N V V I Q P R N K H S S Q C A
 26761 - CCTCTCCGGGGACATTGTGACCCAGACCCGTCATGGAAAGTGAACTTGTCAATGGTCT - 26820
 - P L R G T I V T R P L M E S E L V I G A
 - L S G G Q L * P D R S W K V N L S L V L
 - S P G D N C D Q T A H G K * T C H W C C
 26821 - GTGATCATCTGCTGACTTCGAATGCCGAGACACTCCCTAGGGCGTGTGACATTAG - 26880
 - V I I R G H L R M A G H S L G R C D I K
 - * S F V V T C E W P D T P * G A V T L R
 - D H S W S L A N G R T L P R A L * H A G

FIG. 11 Con't

51/106

26881 - GACCTGCCAAAGAGATCAGTGGCTACATCACGAGCTTCCTTATTACAATTAGGA - 26940
 - D L P K E I T V A T S R T L S Y Y K L G
 - T C Q K R S L W L H H E R F L I T N * E
 - P A K R D H C G Y I T N A F L L Q I R S
 26941 - GGGTCCGAGCGTGAGGCACGTGATTCAGGTTTCTGCTCATCACANCCGCTACCGTATTGGA - 27000
 - A S Q R V G T D S G F A R A Y N R Y R I G
 - R R S V * A L I Q V L H T T A T V L E
 - V A A C R H * F R F C C I Q P L P Y W K
 27001 - AACTATAAATTAAATACAGACCAAGCGGGTGCAGCAGCAATTAGCTTCTAGTACAG - 27060
 - N Y K L N T D H A G S N D N I A L L V Q
 - T I N * I Q T T P V A T T I L L C * Y S
 - L * I K Y P R R * Q R Q Y C F A S T V
 27061 - TAGTGACACAGATGTTCTGACTTCAGGTTAACATAGCAGAGATATTGAT - 27120
 - V T T D V S S C * L F G Y N S R D I D
 - K * Q Q M F H L V D F Q V T I A E I L I
 - S D N R C F I D L T S R L Q * Q R Y * L
 27121 - TATCATTAGGAGGACTTCAGGATGCTATTGGRAATCTGACSTTAAATAAGTTCTAT - 27180
 - Y H Y E D F O D C Y C L E S * R Y N K F N
 - I M R T F R I A I W N L D V I I S S I
 - S L * G L S G L L F G I L T L * * V Q *
 27181 - AGTGAGACRATTATTATGCTCTRAACTAAGAGAATTATTCGGAGTTAGATGATGAAGA - 27240
 - S E T I I * A S N * E E L F G V R * * R
 - V R Q L F K P L T K K N Y S E L D D D E
 - * D N Y L S E L * L R R I I R S * M M K N
 27241 - ACCTAAGGAGTAGGATTACCATAAACAGACATGAAAATTATTCCTCTGACATGA - 27300
 - T Y G V R L S I K R T * K L F S S * H *
 - P M E L D Y P * N E H E N Y S L P D I D
 - L W S * I I H K T N N M K I I L F L T L I
 27301 - TTGTATTACATCTTGCAGGCTATACTACTATCAGGAGTGTGTTAGAGGTACGACTGTAC - 27360
 - L Y L H L A S Y I T I R S V L E V R L Y
 - C I Y I L R A I S L S G V C * R Y D D C T
 - V F T S C E L Y H Y Q E C V R G T T V L
 27361 - TACTAAAGAACCTTGCCCATCAGGAACATACGGGCAATTACCCATTCTACCCCTCTG - 27420
 - Y * K N L A H Q E H T R A I H H F T L L
 - T K R T L P I R N I R G O F T I S P S C
 - L K E P C P C S G T Y E G N S F H F P L A
 27421 - CTGACAAATTGAGACTAACATTGCACTAGCACACACTTGCCTTCTGCTGAGC - 27480
 - L T I N L H * L A L A H T L L L L V L T
 - * Q * I C T N L H * H T L C F C L C * R
 - D N K F A L T C T S T H F A F A C A D G
 27481 - GTACTCGACATCCTTACGTGCTGGCAAGATCATTTCCACAAACATTTCATCAGAC - 27540
 - V L D I P I S C V Q D Q F H Q N F S S D
 - Y S T Y L S A R A C K I S F T K T F H Q T
 - T R H T Y Q L R A R S V S P K L F I R Q
 27541 - AAGAGGAGGTTCAACAGAGCTACTCGGCACTTTCATGTCCTCTAGTAT - 27600
 - K R R F N K S S T R F F S S L L L L * Y
 - R G G S T R A L L A T F S H C C C C S S I
 - E E V Q Q E L Y S P L F L I V A A L V F
 27601 - TTTTAATACTTTGCTCACCATTAAGAGAAAGACAGAATGAATGAGCTCACTTTAATGA - 27660
 - F * Y F R S P L R E R Q N E * A E F N *
 - F N T L L H H * E K D R M N E L T L I D
 - L I L C F T I K R K T E * M S S L * L T
 27661 - CCTCTATTGTGCTTTAGGCTTCTGCTATTCCTGTTAAATGCTTATTATAT - 27720
 - L L F V L F S L S A I P C F N N A Y Y I
 - F Y L C F L A F L L F L V L I M L I I F
 - S I C A F * P F C Y S L F * + C L L Y F

FIG. 11 Con't

52/106

27721 - TTGGTTTCCTCGAAATCCAGGATCTAGAAGAACCTTGACCAAAAGTCAAACGACAT - 27780
 - L V F T R N P G S R R T L Y Q S L N E H
 - W F S L E I Q D L E E P C T K V * T N M
 - G F H S K S R I * K N L V P K S K R T *
 27781 - GAAACTCTCATGGTTGACTGATTCTCTATGCAGTGACTGATAGCAGCTGTAGTACA - 27840
 - E T S H C F D L Y F T S M Q L H M H C S T
 - K L L I V L T C I S L C S C I C T V V Q
 - N F S L F * L V F L Y R V A Y A L * Y S
 27841 - GCGCTGTGCTCAATAAACCTCATGCTGAGATCCTGTAGGTCACRACACTAGGG - 27900
 - A L C I * * T S C A * R S L * G T T L G
 - R C A S N K P H V L E D P C K V Q H * G
 - A V H L I N L K I L V R Y N T R G
 27901 - GTAATACTAGACACTGCTGGCTTGTGCTCAGGAAGGTTTACCTTTATAGAT - 27960
 - V I L I A L L G F V L * E R F Y L F I D
 - * Y L * H C L A L C S R K G F T F S * M
 - N T Y S T A W L C A L G K V L P F H R W
 27961 - GGACRACTAGGTCACACGACACCTAATGTTACTATCAGTCAGATCAGCTG - 28020
 - G T L W F K H A H L M L L S T V K I Q L
 - A H Y G S N M H T * C Y Y Q L S R S S W
 - H T M V Q T C T F N V T I N C Q D P A G
 28021 - GTGGTGCCTTAATAGCTAGGGTGTGGTACCCCTCATGAGGTACCCAACTGGCTGATT - 28080
 - V V R L * L G V G T F M K V T K L L H L
 - W C A Y S * V L V P S * R S P N C C I *
 - G A L I A R C W Y L H E G H Q T A A F R
 28081 - GAGACGACTGTGTGTTAAATAGCAGCACAAATTAATGTCGATATGGACCCCCA - 28140
 - E T Y L L F * I N E Q I K M S D N G P Q
 - R R T C C F K * T N K L K C L I M D P N
 - D V L V V L N K R T N * N V * * W T P I
 28141 - TCAAAACRNACGTAGTAGCCCGCCGATTACATTGGTGGACCCACAGATTAACGTACAT - 28200
 - S N Q R S A P R I T F G G F T D S T D N
 - Q T N V V P P A L H L W D P Q I Q L T I
 - K P T * C P P H Y I W T H R F N * Q *
 28201 - AACCGAAGATGGAGGCGCRAATGGGCCRAAGGCCAAAACAGGGCCGQACCCCAAGGTTTACCC - 28260
 - N Q N G N G G A R P K Q R R P Q G L P
 - T R M E D A M G Q G O N S A D P K V Y P
 - P E W R T Q W G K A K T A P T P R F T Q
 28261 - AATAAATCTGCGCTTGTGTCACAGCTCACTCAGCATGGCAAGGGGAACTTAGTTTC - 28320
 - N N T A S W F T A L T Q H G K E E L R F
 - I I L R L G S Q O L S L S M A R R N L D S
 - * Y C V L W H S S H S A W Q G G T * I P
 28321 - CCTCGAGGGCAGGGCGTTCCATCACACCAATAGTGGTCCAGATGGACCCAAATTGGCTAC - 28380
 - P R G Q G V P I N T N S G P D D Q I G Y
 - L E A R A F Q O S T P I V V Q M T K L A T
 - S R F G R S N O H Q * W S R * P N W L L
 28381 - TACCGAAGAGCTACCCAGAGTGTGGTGTGAGCAGGCAAAATGAAAAGACTCAGCCCC - 28440
 - Y R R A T R R V R G G D G K M K E L S P
 - T E E L P D E F V V V V T A K * K S S A P
 - P K S Y P T S S W W * R Q N E R A Q P Q
 28441 - AGATGGTACTCTTATTACCGTGGAACTGGCCCAAGAGCTTCACCTCCCTACGGCCCTAAC - 28500
 - R W Y G Y L G T G P E A S L P Y G A N
 - D G T S I T * E L A Q K L H F P T A L T
 - M V L L L P R N W P R S F T S L R R * Q
 28501 - AAAGAAGGCATGTGTTGCAACTGGGGAGCTGGCCCAAGAGCTTCACCTCCCTACGGCCCTAAC - 28560
 - K E G I V W V A T E G A L N T P K D H I
 - K K A S Y G L Q L R E P * I H P K T T L
 - R R H R M G C N * G S L E Y T Q R P H W

FIG. 11 Con't

53/106

28561 - GGCACCCGCGAATCCTAATAAACATGCTGCCRCGGCTACAACCTTCCCTCAGGAAACRACA - 28620
 - G T R N P N N H A A T V I Q L P Q G T T
 - A P A I L I Y T M L P P C Y N F L K E Q H
 - H P Q S * + Q C C H R A T T S S R N N I
 - 28621 - H P Q S * + Q C C H R A T T S S R N N I
 - T T G C C A A A G G C T T C T G C G A G G G G A C G A G G G C G C A G T C A A G C C T C T C G C T C C - 28680
 - L P K G F Y A E G S R G G S Q A S S R S
 - C Q K K A S T Q R E R A E A A V K P L L A P
 - A K R L L R R G K Q R Q S S L F S L L
 - 28681 - T C A T C A C G T A G T C G C G G T A A T C A G A A A T T C A A C T C C T G G C A G C A T G G G G / A A T T C T - 28740
 - S S R S R G N S R N S R N S T P G S S R G N S
 - H H V V A V I Q E I Q L L A A V G E I L
 - I T * S R F K F N S W W Q Q * G K F S
 - 28741 - C C T G C T C G A T G G C T A G C G G A G G T G G G A A C T G C C C T C G C C T A T T G C T G C T A G C A G A - 28800
 - P A R M A S G G G E T R A L L L L D R
 - L L E W L R E V V K L P S R Y C C * T D
 - C S N G * R R R W * N C P R A I A A R Q I
 - 28801 - T T G A C C G C T T G G A G G C R A A G T T C T G G T R A A G G C C A R A C A R A C A R G G C C A A R C T G T C - 28860
 - L N Q L E S K V S G K G Q Q Q Q Q G Q T V
 - * T S L R A K F L V L K A N N N N K A K L S
 - E P A * E Q S F W * R P T T T T R P N C H
 - 28861 - A C T A A G A A A T C T G C T G C T G A G G C A T C T A A A A G C T C G C C A A A A A C C T A C T G C C A C A A A A - 28920
 - T K K S A E A S K P R Q K R T A T K
 - L R N L L L R H L K S L A K N V L P Q N
 - * E I C C * G I * K A S P K T Y C H K T
 - 28921 - C R A G T A C A C T C A C T C R A G C A T T T G G G A G A C O T G G T C C A G A C A A C C C A A G G A A R A T T C - 28980
 - Q Y N V T Q R F G R R G P E Q T Q G N F
 - S T T S L K H L G D V V V Q N K P K E I S
 - V Q R H S S I V T W S R T N P R K F R
 - 28981 - G G G A C C A A G R C T T A A T C A G C A A C G G A C T G A T T A C A A C A T T G G C C C A A T T G C C A A - 29040
 - G D Q D L I R Q G T D Y K H W P Q I A Q
 - G T K T * S D K E L I T N I G R K L H N
 - G P R P N Q T R N * L Q T L A A N C T I
 - 29041 - T T T G C T C C A A G T G C C T C G C A T T C T T G G A A T G T C A G G C A T T G G C A T G G A A G T C A C A C C T - 29100
 - F A P S A S A F F G M S R I G M E V T P
 - L L Q V P L H S L E C H A L A L A W K S H L
 - C S K C L C I L W N V T H W H G S H T F
 - 29101 - T C G G G A A C A T T G G T G A C T T A T C A T G A G G C C A T T A A T T G G A T G C A A A G A T C C A A A T T C - 29160
 - S G T W L T Y H G A I K L D D K D P Q F
 - R E H G * L I M E P D L N W M T K I H N S
 - G N M A D L S W S H * I G * Q R S T I Q
 - 29161 - A A A G A C A C C T C A T C T G C A C A G C C A C T G A C C C A T A C A A A C A T T G C C C A C A C A - 29220
 - K D N V I L L N K H I D A Y K T F P P T
 - K T T T S Y C * T S T L T H T K H S H Q Q
 - R Q R H T A E Q A H * R I Q N I F T N R
 - 29221 - G A G G C T A A J A A G G J A C A A M J G A A A M G C T G A T G A G C T C A G C C T T T G C C C A G A G A C A - 29280
 - E P K K C D K K K K T D E A Q P L P Q R Q
 - S L K R T K R K R L M K L S L C R R D K
 - A * K G Q K E K D * * S S A F A A E T K
 - 29281 - A A G A A G C A G C C C A C T G T G A C T C T T C T C G C G G C T G A C A T G C A T G A T T T C T C A G R C A A - 29340
 - K K Q P T V T L L P A R D M D D F S R Q
 - R S S P L * L F F F L R L T W M I S P D N
 - E A A H C D S S S C G * R G * F L Q O T T
 - 29341 - C T T C R A A A T T C A T G A T G G A G G C T C T G C T G A T T C A A C T C A G G C T A A A C A C T C T G A T G - 29400
 - L Q N S M S G A S A D S T Q A * T L M M M
 - F K I P * V E L L L I Q L R H K H S * *
 - S K F R E W S F C * F N S G I N T H D D

FIG. 11 Con't

54/106

29401 - ACCACACAAGGGCAGATGGGCTATGTAAACGTTTCGAATTCCGTTACGATACATAGTC - 29460
 - T T Q G R W A M * T F S Q F R L R Y I V
 - P H K A D G L C K R F R N S V Y D T * S
 - H T R Q M G Y V N V F A I P F T I H S L
 29461 - TACTCTTGTGCAAGAATGAATTCTCGTACTAACAGCACRAGTAGGTTAGTTAACITTA - 29520
 - Y S C A E * I L V T K O H K * V * L T L
 - T L V Q N E F S * L N S T S R F S * L *
 - L L C R M N S R N * T A Q V G L V N F N
 29521 - ATCTCACATAGCAATCTTATCATGTCATTTAGGGAGACTTGAAGAGCCACCA - 29580
 - I S H S N L * S M C N I R E D L K E P P
 - S H I A I F N Q C V T L G R T * K S H H
 - L T * Q S L I N V * H * G G L E R A T T
 29581 - CATTTTCATCGAGGCCACGCCGAGTAGCATCGAGGGTACAGTGATAATGCTAGGGAG - 29640
 - H F H R G H A E Y D R G Y S E * C * G E
 - I F I E A T R S T I E G T V N N A R E S
 - F S S R P R G V R S R V Q * I M L G R A
 29641 - CTGCTATATGGAGAGCCCTAAATGTTAAATTATTTAGTAGTGCTATCCCCATGTG - 29700
 - L P I W K S P N V * N * F * * C Y P H V
 - C L Y G R A L M C K I N F S S A I F M *
 - A Y M E E P * C V K L I L V V L S S P C D
 29701 - ATTTAAATAGCITCTTCTAGGAGAAATGACAAAAAAAAAAAAAAA - 29742
 - I L I A S * E N D K K K K K X
 - F * * L L R R M T K K K K X
 - F N S F L G E * Q K K K K X

FIG. 11 Con't

55/106

1 - TTTTTTTTTTTTGTCAATTCTCTAAAGAGCTTAAATCAGGGGATAGCACTA - 60
 - F F F F F L S F S * E A I K I T W G * H Y
 - F F F F C H S P K K L L K S R G D S T T
 61 - CTAAAATTAATTTCACATTAGGGCTTCCATATAGCACGCTCCCTAGCATTATTC - 120
 - L K L I L H I R A L P Y R O S L S A L F
 - * N * F Y T L G L F H I G S S P * H Y S
 - K I N F T H * G S S I * R A L P S I I H
 121 - ACTGTACCTCGATCGTACTCCCGTGGCTCGATGAAATTGGTGGCTTTCARGTC - 180
 - T V P S I V L R V A S M R M W W L F Q V
 - L Y P R S Y S A W P R * K C G G S F K S
 - C T L D R T P R G L D E N V V V A L S S P
 181 - CTCCCTAACTTACATTAGATAAGATGCTATGAGATAAAGTAACTAACCTA - 240
 - L P N V T H * L K I A M * D * S * L N L
 - S L M L H I D * R L L C E I K V N * T Y
 - P * C Y T L I K D C Y V R L K L T K P T
 241 - CTTGGCTTGTAGTAGGAGATTCACTGCACRAAGTAGACTATGATATCGAAACG - 300
 - L V L F F S Y E N S F C T R V D Y V S * T
 - L C C C L V T R I H S A Q E * T M Y R K R
 - C A V * L R E F I L H K S R L C I V N G
 301 - GAAATGGGAAACGTTACATAGCCCATGCCTGGTGGTCATCATGAGTTATG - 360
 - E L R K R L H S P S A L C G H H E C L C
 - N C E N N V Y I A H L P C V V I M S V Y R
 - I A K T F T * P I C L V W N S S * V F M P
 361 - CTGAGTGAACTAGCGAGAGCTCACTCATGGATTGAGTTGCTGGAGAATCATC - 420
 - L S * I S R S S T R H G I L K L S G E I I
 - * V E S A E A P L M E F * S C L E K S S
 - E L N Q Q K L H S W F E V V V N R N H P
 421 - CATGTCAGCGCAGGAAGAAAGAGTCACAGTGGGCTGCTTGTCTCGCGCAAAGG - 480
 - H V S R R R K K S H S G L L L S L R Q R
 - M S A A G R R R V T V G C C F F C L C G K G
 - C Q P Q E E R E S Q W A A F V S A A K A
 481 - CTGAGCTCATCAGTTCTTCTTGTCTTGTAGCTCTGTGGGGATGTTT - 540
 - L S F I S L F L V F R L C W N E C F
 - * A S S V F F F L S F L G S V G G N V L
 - E L H Q S F S F C P F * A L L V G M F C
 541 - GATGCGTAACTGTCCTGTCAGCACTATGAGCTGTCTTGTGATGATCTGTC - 600
 - V C V L W L Q Q Y D V V F E L W I E V
 - Y A S M C L F S S M T L S L N C G S L S
 - M R Q C A C S A V * R C L * I V D L C H
 601 - ATCCAATTATGGCTCATGATAAGTCAGCCATGGTCCCGAAGGTGACTCTGCC - 660
 - I Q F N G S M I S Q P C S R R C D F H A
 - S N L M A P * V S H V P E G V T S S M P
 - P I * W L H D K S A M F P K V * L P C O
 661 - AATGGGYGRCATCCAGAATGCAAGGGCACTGGGCCAATTGTCATTTGGGCCA - 720
 - N A * H S K E C R G T W S K L C N L R P
 - M R D I P K N A E A L G A N C A I C G Q
 - C V T F Q R M Q R H L E Q I V Q F A A N
 721 - AIGTTTGTAATCAGTTCTGTCGATAGGTCTGACTCTGGGGAAATTCCCTGGGTG - 780
 - M F V I S S L S D * V L V P E I S L G L
 - C L * S V P C L I R S W S P K F P W V C
 - V C N Q P L V * L G L G P R N F L G F V
 781 - TTCTGGRCACGTCCTCCAAATGCTGAGTGCAGCTGACTGTTTGTGGCAGCTTT - 840
 - F W T T S P K C L S D V V L F C G S T F
 - S G P R L P N A * V T L Y C F V A V R F
 - L D H V S Q M L E * R C T V L W Q Y V F

FIG. 12

56/106

841 - TTGGCGAGGCCTTTAGATGCTCAGCAGCAGATTCTTAGTGCAGTTGGCCCTGGT - 900
 - L A R L F R C L S S R F L S D S D L A L L
 - W R G F L D D A S A A D F L V T V W P C C
 - G E A F * M P Q Q Q I S * * Q F G L V V
 901 - TTGTTGGCCTTAACAGAAACTTGTCTCAAGCTGTTCAATCTGCTAGCACCAATAG - 960
 - L L A F T R N F T L A L K L V Q S V * Q Q *
 - C W P L P E T N F T L S S W F N L S S S N S
 - V G L Y Q K L C S Q A G S I C L A A I A
 961 - CGCGAGGCCAGTTTACCAACCTCCGCTAGCCATTCGAGCAGGAGAATTCCCTACTGCT - 1020
 - R E G S F T T S A S H S S R R I S P T A
 - A R A V S P P P L A I R A G E F P L L L
 - R G Q F H H L R * P F E Q E N F F P Y C C
 1021 - GCCAGGACTGTTATCTGAATTACCCGACTACCTGATGAGGAGCAGAGGGCTG - 1080
 - A R S * I S * I T A T T * * G A R R G L
 - P G V E F L E L P R L R D E E R E E A *
 - Q E L N F L D N Y R D Y V M R S E K R L D
 1081 - ACTGCCGCTCTGCCTCTCGCTAGAACCTTGGCAATGTTCTTGAGGAAG - 1140
 - T A A S A S L C V E A F W Q C C S L R K
 - L P P L L P S A * K P F G N V V P * G S
 - C R L C F P L R R S L D L A M L F L E E V
 1141 - TTGTAGCACGGTGGCAGCATTTGTTATTAGATTGGGGTGCCTAATGTGGCTTTGGGTG - 1200
 - L * H G G S I V V I R I A G A N V V F G C
 - C S T V A A L L G L R V P M W S L G V
 - V A R W Q H C Y * D C G C Q C G L W V Y
 1201 - ATTCAAGCTCCCTCAGTGCACCCCATACGATGCCCTTCTGGTAGGCCGTAGGGAG - 1260
 - I Q G S L S C N P Y D A F F F V S A V G K
 - F K A P S V A T H T M P S L L A P * G S
 - S R L P Q L O P I R C L L C * R R R E V
 1261 - TGAAGCTCTGGGCCAGTCTCTAGTAATAGAAGTACCATCTGGGCCTGAGCCTCTTCAT - 1320
 - * S F W R A S S * V I E V P S G A E L F H
 - E A S G P V P R * * K Y H L G L S S F I
 - K L L G Q F G L G N R S T I W G * A L S F
 1321 - TTGCGCTGACCAACCCGAACTCTGGGTAGCTCTGGGTAGTACGCCATTGGTCATC - 1380
 - F A V T T N S S G S S S V V A N L V I
 - L P S P P P R T R R R V A L R * * P I W S S
 - C R H H E L V G * L F G S S Q F G H L
 1381 - TGGACCACTATTGGTTGTGATGGACCCCTGGCTGGGAACTTAAGTCTCTCCCT - 1440
 - W T T I G Q V D W N N A L A S R E S K F I L L
 - G P L L V L L I G T P W P R G N L S S S L
 - D H Y W C * L E R P G L E G I * V P P P C
 1441 - GCCATGCTGACTGAGAGCTGAGAACCAAGCAGCTATTATGGGTAACCTGGGGTC - 1500
 - A M L S E S C E P R S I I G * T L G S
 - P C Y V R A V N Q D A V L L G K P W G R
 - H A E * E L * T K T Q Y Y W V N L G V G
 1501 - GCGCTGTTTGGCCTTGCCCCATTGGCTCCCATCTGGTATATTGTCAGTGAATCTGT - 1560
 - A L F W P C P I A S S I L V I V S * I C
 - R C F G L A P L R P P F W L L S V E S V
 - A V L A L F H C V L B S G Y C Q L N L W
 1561 - GGGTCCACCAATGTAATGGGGGGCACTACGGTTGGTTGATTTGGGTCCATTATCGA - 1620
 - G S T K C N A G G T T L V * L G S I I R
 - G P P N V M R G A L R W F D W G P L S D
 - V H Q M * C G G H Y V G L I G V H Y Q T
 1621 - CATTAAATTGTTGCTTATTAAACAAAGTACGCTCTAAATGCAGCAGTTGGT - 1680
 - H F N L F V Y L K Q Q V R L * M Q Q F G
 - I L I C S F I * N N K Y V S K C S S L V
 - F * F V R L F K T T S T S L N A A V H *

FIG. 12 Con't

57/106

1681 - GACCTTCATGAAAGGTACCAACACCTAGCTATAAGCGCACCACCGCTGGATCTTGACAGT - 1740
 - D L H E G T N T * L * R H H Q L D L D S
 - T F M K V P T P S Y K R T T S W I L T V
 - P S * R Y Q H L A I S A P P A G S * Q L
 1741 - GATGATGTAACATTAGGTGTCATGGTGAACCATAGTGTGCCATCTATGAAAAGGTAAAA - 1800
 - * * * H * V C M F E P * C A I Y E K V K
 - D S N I R C A C L C L N H S V P S M K R * N
 - I V T L G V H V * T I V C H L * K G K T
 1801 - CCTTTCCATAGAGCACRAACCCAAAGCAGTGTCTATAAGTATTACCCCTAGTGTGACCTTA - 1860
 - P F F L E R H K A K Q C Y K Y Y P * C C T L
 - L S * S T K P S S A I S I T P S V V P Y
 - F P R A Q S Q A V L * V L P L V L Y L T
 1861 - CAAGGATCTCAACGCCATAGGGTTTATTAGATGTCACGCCGTACTACAGTCATATG - 1920
 - Q G S S S T * G L L D A Q R C T T V H M
 - K D L Q A H E V Y * M H S A V L Q C I C
 - R I F K H M H R F I R C T A L Y Y S A Y A
 1921 - CAACTGCATAGAGAAATACAGTCRAACATGAGRAGTTTCATGTTGCTTTAGRCTTGG - 1980
 - Q L H R E I Q V K T M T R S F M F V * T L
 - N C I E K Y K S Q * E V S C S F R L W
 - T A * R N T S Q N N E K F H V R L D F G
 1981 - GTACAAGGATCTCTCTAGATCTGGATTGAGATGAAAACCAAATAATAAGCATTATT - 2040
 - V Q G S S R S W I S S E N Q N I I S I I
 - Y K V L L D P G F R V K T K I * * A L L
 - T R F * I L D F E * K F K Y N K H Y *
 2041 - AAAACAAAGGAATAGCAGAAAGCTAAAAGCACAAATAGAGTCATTAAGTCCTCA - 2100
 - K T R N S R K A K K H K * K S I K V S S
 - K Q G I A E R L K S T N R S Q L K * A H
 - N K E * Q K G * K A Q I E V N * S E L I
 2101 - TTCAATTCTGCTTCCTTAATGGTAGAGCAAGGTTAAAATACTAGAGCAGCACAA - 2160
 - F I L S F S * W * S K V L K I L E Q Q O
 - S F C L S L N G E A K Y * K Y * S S N N
 - H S V F L L M V K Q S I K N T R A A T M
 2161 - TGAGAAGAAATGGCCGAGTAGGGCTCTGTTGACCTCCCTTGCTGTGAAAGTTTG - 2220
 - * E K V A S R A L V E P L V * * K V L
 - E K K W R V E V L L N L L L S D E K F W
 - R K S G E * S S C * T S S C L M K S F G
 2221 - GTGAAACTGATCTGCRGCGCTGATGGTAGCTCGAGTACCGTCACCCACARGCRAAG - 2280
 - V K L I L H A D R Y V E Y R Q H M K Q K
 - * N * S C T Q L D I G M S S T V S T S K S
 - E T D L A R S * V C R V P S A Q A K A
 2281 - CAAAGTGTGCTCTAGTCAGATTAGTCGAATTATTTGTCAGCAAGGGTGAATGGTG - 2340
 - Q S V C * C K L V Q I Y C Q O E G E M V
 - K V C V A S A S * C K F I V L S K R V K W *
 - K C V L V Q V S A N L L S A R G * N G E
 2341 - AATTCGCCCTGCTATGTCCTGATGGCAAGGTTCTTAACTAGTACAGTCGTACCTCTAA - 2400
 - N C P R M F L M G K V L L V V O S Y L *
 - I A L V C S * W A R F F * * Y S R T T S N
 - L P S Y V P D G Q G S F S S T V V P L T
 2401 - CACACTCTGATAGTGAATAGCTGGCAAGATGAAATACATCATGTCAGGAAAGAGAA - 2460
 - H T P D S D I A R K M * I O S M S G R E
 - T L L I V I * L A R C K Y N Q C Q E E N
 - H S * * * Y S S Q D V N T I N V R K R I
 2461 - TAATTTCTATGTCGTTTATGGATAATCTAACCTCCATAGGTCTTCATCATCTAACCTC - 2520
 - * F S C S F Y G * S N S I G S S S S N S
 - N F H V R F M D N L T P * V L H H L T P
 - I F M F V L W I I * L H R F F I I * L R

FIG. 12 Con't

58/106

2521 - G R A T A T T C T C T T A G G G C T T A A A T T G T C T C A C T T G A A C T T A T T A T A C G - 2580
 - E * F F L V R G L N N C L T I E L I I T
 - N N S S * L E A A * I I V V S L L N L L * R
 - I I L L S * R L K * L S H Y * T Y Y N V
 2581 - T C A R G A T T C C A A T T A G C A A T C C T G A A A G T C C P C T A A T G T A A T C A A T A T C T G C T A T T - 2640
 - S R F Q I A I I L K V L I M I I N I S A I
 - Q D S K * Q S * K S S * * S I S L L L
 - K I P N S N P E S P H N D N Q Y L C Y C
 2641 - G T A A C C T G S A A G T C A C A A G T G A A A C A T C T G T G C T A C T A C T G T A C T A G C A A G C C A T - 2700
 - V T W K S T R * N I C C H L L Y * Q S N
 - * P G S Q Q D E T S V T Y C T S K A I
 - N L E V N K M K H L L S L T V L A K Q Y
 2701 - A T T G T C G T G C T A C C G G C T G G T C I G T A T T T A T T A T G T T C C A A T C G G T A G C G G T T - 2760
 - I V V A T G V V C I * P I V S N T V A V
 - L S L L P P A W S V F N L * F P I R * R L
 - C R C Y R R G L L I Y S F Q Y G S G C
 2761 - G T A T G C A G C R A A A C C T G A A T C A G T G C C T A C A C G C T G C A G C G C T C T A A T T G T A A T A G A - 2820
 - V C S K T * I S A Y T L R R S * F V I R
 - Y A A K P E S V P T R C D A P N L * * E
 - M Q Q N Q C L H A R A T L L I C N K K
 2821 - A A G C G T T C G T G A T G A G C C A C A G T G A T C T C T T G G C A G G T C T I A A T G T C A C G G C C C - 2880
 - K R S * C S H S D L F W Q V L N V T A P
 - S V R D V A T V I S F G R S L M S Q R P
 - A F V M * P Q * S L L A G P * C H S A L
 2881 - T A G G G A G T G T C C G C A T T C G C A G T G A C C G A C G A T G A T C A C A G C A C C A T G A C A A G T C - 2940
 - * G V S G H S Q V T T N D H S T N D K F
 - R E C P A I R K * P R M I T A P M T S S
 - G S V R P F A S D H E * S Q G * Q V H
 2941 - A C T T T C C A T G A G C G C G T C G C A C A T T G T G C C C G G A G G G C A C A T T G A G A G A G A T G T T - 3000
 - T F H E R S G H N C C P P E R H I E K N V
 - L S M S G L V T I V P R R G T L R R M F
 - F P * A V W S Q S L P S G E A H * E E C L
 3001 - T G T T T C G G G T T G A T G A C C A C A T T G A G G G G T A G A G G C M A A C A G G C T G A A G G A G C A M C - 3060
 - C F W V E * P H * A G T S K Q P E G S N
 - V S G L N D H I E R V R A N S L K E A T
 - F L G * M T L T L S G Y E Q T A * R K Q R
 3061 - G A A G T A C T G A C C A C A T C A G C C T A C A T C A G C C A T T G C A T C G C A T C C G C A G T - 3120
 - E V A K P H Q K A P T I Q A I A I A I A I P P V
 - S S * A T T S L Q Y K P L Q S Q S R Q S
 3121 - C A C C C A A T T A T T C T G T A G A C A C A G C A G C A C A A A C A A G C A A G T G T T A C T G G C C A C A - 3180
 - H P I N S V D N S K H K T S K C Y W P Q
 - T Q L I L * T T A S T K Q A S V T G H K
 - P N * F C R Q Q Q Q A Q N K Q V L L A T R
 3181 - G A G C C A G A G G A A A C A R A G C T T T T A T G T A C A A A A C C T G T T C G A T T G A T A T G G C A A A - 3240
 - E P E E N E L Y Y V Q K P V P R I G K
 - S Q R K T S F I M Y K N L F R L E * A N
 - A R G K Q A L L C T K T C S D * N R Q I
 3241 - T T G T A G T A C A T A T C C A G G C T A G G A T A G G A A A C C T A T T A C T A G G T T C C A T T G T T C A G - 3300
 - L * * H N P G * E * E T Y Y * V P L F Q
 - C S N E I Q A R R N R K P I T R F H C S R
 - V V T * S R L G I G N L L L G S I V P G
 3301 - G A G T T G T T T A A G C T C T C A A C G G T A A T G T A C C G T T G C T G C A T G A T A A G C A A T G T T A A - 3360
 - E L F K L L N G N S T V V C H D K Q C *
 - S C L S S S T V I V P L S A M I S N V K
 - V V * A P Q R * * Y R C L P * * A M L K

FIG. 12 Con't

59/106

3361 - AGTTCCAACAGAATAATAATAATAGTTAGTCGTTAGACAGAAAGATCAGGAACCTCT - 3420
 - S S K Q N N N N S * F V * T R R S G T P
 - V P N R I I I I I V S S F R P E D Q E L L
 - F Q T E * * * L V R L D Q K I R N S F
 3421 - TCAGAAGAGTTCAAGATTITTAACCGCGAGTAGACGTAAACCGTTGGTTTACTAAACTC - 3480
 - S E E F R F L T R E * T * T V G F T K L
 - Q K S S D F * H A S R R K P L V L L N S
 - R R V Q I F N T R V D V N R W F Y * T H
 3481 - ACGTTAACATATTGAGCAGTAGCGCACACRATCGAAGCCAGTAAGGATGGCTAGTGTG - 3540
 - T L T I L Q Q Y A H N R S A V R M A S V
 - R * Q Y C S S T H T I E A Q * G W L V *
 - V N N I A A V R T Q S K R S K D G * C D
 3541 - ACTAGGAAGATAACCGAAGCAAGAAAAGAAGTAGCGCTATTAACTATTAAAGTACCT - 3600
 - T S K N T T K A R K R S T L L T I N V P
 - L A R I P R K Q E K E V R Y * L L T Y L
 - * Q E Y H E S S K K K Y A I N Y * R T C
 3601 - GTTCTTCGGAACGAAATGAGTACATAAGTTCTGACTCTCTTGCTCTAACAAAGGC - 3660
 - V S S E T T N E Y I S S Y S C L S C A Y K G
 - F L P K R M S T * V R T R H F L V L T K A
 - F F R N E * V H K F V L T F L C L Q R H
 3661 - ACGCTAGTAGTCGTCGTCGCGCTCATCATAAARTGGATCATTGCTGGATTAGCAACTCT - 3720
 - T L V V V V G G S * I G S I A G L A T P
 - R * * S S S A H H K L D P L L D * Q L D
 - A S R R R L I I N H C W I S N S *
 3721 - GAAGAGCCGCTGGATGGTGTGATTTGACATTCCGGTGGCTTTAACAGCTTGTAAAG - 3780
 - E E P S I V C I C T F G G S L T S L L K
 - K S R R L C V F A H S V G L * Q A C * R
 - R A V D C V Y L H I R W V F N K L V K D
 3781 - ATGAAGATAATGAGCTTTCAATACCAAGTGTCTGATTAATTGTGTAGACTCAAGCTG - 3840
 - M K N V A F S I P V S V V I C V D S S W
 - * R M * H F Q Y Q C L D * * F V * T Q A G
 - E E C S I F N T S V C S N L C R L K L V
 3841 - TAGTAACTCTGGTGAATAGGCATAGTACAACGACATAGCTTTAACACCTGAGTCGCTA - 3900
 - * * T S V K * P C T T T * S L T P E C L
 - S K L R * N I S H V Q R H S L * H L S A Y
 - V N F G E I A M Y N D I V F N T * V P I
 3901 - TCTCTAGAAATAACCAAATTGGTAGTCCTCTTGGATTTGGTGTGAAATGCCGTCA - 3960
 - S S E * P P I W * S S L S F G V E M P S
 - P Q N N H Q F G S S L L * V L V L K C R H
 - L R I T T N L V V F F E F F W C * N A V T
 3961 - CCTTCAGTAGCAAAATGTGATCTGAGACTGTATAGGTATAACGATAGTCATAGTTA - 4020
 - P S V T T I V S V T L L Y G I Q * S * L
 - L Q * R Q L Y L * H C Y I M V Y S S H S Y
 - F S N D N C I C D T V I W Y T V V I V M
 4021 - TGTTGTCGCCAGCAACAAAGTAGTTGCACTATAAGTAATGGGTTCTGGATTTCAC - 4080
 - C V C Q Q T K * L A S * S N G F L D L H
 - V C A S K O S S W H H K V M G S W I C T
 - C V P A N K V V G I I K * W V L G F A L
 4081 - TTCCAAACAAAGCNACTCTCATATAATTCTACATCGGTGATGCACTGTAGAAAATAT - 4140
 - F Q Q S Q H L I I L H A L M H C R K Y
 - S N K A N I S * * F Y M R * C I V E N I
 - P T K P T S H N N S T C V D A L * K I Y
 4141 - ATCAGGCATAGAGTACAAAAATTGGCGCTCTTACCTGACGGACAAAGATCT - 4200
 - I K A * R Y K N C A S L P F A A T S K R C
 - S R H R G T K I A P P Y L Q R Q A K D V
 - Q G I E V Q K L R L L T C S D K Q K M *

FIG. 12 Con't

60/106

4201 - GAATAGATGGTAAACAAATAGCAGCGACTAATTGCAAAATGCAACTGGAGGCCCTTATAAAGG - 4260
 - E * M V T N S S S K L Q M N W K P L * R
 - N R W * Q I A A V N C K * T G S F Y K G
 - I D G N K * Q Q * I A N E L E A L I K G
 4261 - GCTAGCGCCATCTTITATTCAGCCCAATTATTTCGTAGCCCTCTGAARARACAGCAGA - 4320
 - A S C H L L D S A I I L V A L * K T A R
 - L A A I F Y * Q O L F W * R S E K Q Q E
 - * D P S F I E R N Y F G S R L K N S K K
 4321 - AATGCAACGCCATAACARCCCATCGGAAAGGGAGTGAGGCTGTAGCGTATCGCTTCT - 4380
 - N A T P I T S H P K G S E A C S G I V A
 - M Q R Q * Q A I R K G V R L V A V S L L
 - C N A N N K P S E R E * G L * R Y R C C
 4381 - GTAGCATGCACTGAGTACTTCGAGGAGAACATTGCAATTTCATGCTGTCAGCTTAATT - 4440
 - V R * T V L A G E B A L S I F T G C A V I
 - * H E Q Y L Q E K H C Q F L L A V Q * L
 - S M N S T C R R S I V N F Y W L C S N *
 4441 - GATCCAACAGATAAATTCATAACAAATCCATAACTGTTATGTGTTATGTATT - 4500
 - D P R V K N L I N K S I S S F M C N V I
 - I Q E * K I S * T N P * V R L C V M * F
 - S K S K K S H K Q I H K F V V V * C N L
 4501 - TGACACCCCTGAGAATGGCTCAGAGTCATCCTCATCAACTTGAGCAAGAACCAAG - 4560
 - * H P * E L A O S H P R O T C S K N H K
 - D T L E N W L R V I L I K L R A R T T R
 - T P L R T G S E S S S N L Q Q E P Q E
 4561 - ACCATGCACCCCTGGAGCRACTCAGCACAACTACTCATGCAACAAAGCAAGATTGTAACCA - 4620
 - S M H P * G N C N N * S C N K A R L * P
 - A C T L E A T A T T S H R A T K Q D C N H
 - H A F L R Q L Q Q L V M Q Q S K I V T M
 4621 - TGACGATGGCAATTGTCAGCAATGAGCGAGCCACATACCAAGGCATTATRAT - 4680
 - * R W Q L V Q * S R A K H T K A I * Y
 - D D G N * S S N E A E P N I P R P F N I
 - T M A I S P A M K P S Q T Y Q G H L I Y
 4681 - ATTGCTCATATTTCACCATCTGAAAGGTCATGACTGATTCATTAAATTTTAGCGG - 4740
 - I A H I F P I L E G Q * V I H L N F * R
 - L L I F S Q F L K V N E * F I * I F S D
 - C S Y F F N S * R S M S D S F K F L A T
 4741 - CCTCATGGCCGGTCATIITCTTGTGAATGTCAGCAGAGCGTTATGCTGAAA - 4800
 - P H * G G Q F L F E C * R Q K R * C L K
 - L I E A V N F F L N V D D R S V N A * N
 - S L R R S I S F * M L T T E A L M P E M
 4801 - TGTCCGCAGATCAACATCGCTGATGATGATTTCGAAGTACTGTCCAGCTCTT - 4860
 - C R Q D Q H L V M Y D F * S T C P A L L
 - V A K I N I W * C M I F E V L V Q L F F
 - S P R S T S G D V * F L K Y L S S S S L
 4861 - TGAATGACTCAGCTCAGCTGAGAGATCATAAACATGTTGTTATGATGCCAAATA - 4920
 - * M S Q A Q V A E D H K L C G C * C Q *
 - E * V K L R L Q R I I N C V V N D A N N
 - N E S S S G C R G S * T V L L M M P I T
 4921 - CGACATCACAAATTCTGAGAGCAATGTATTCGTCAGTAATTATGTCAGAAGA - 4980
 - R H H N F L R Q M Y C L * L F V E K R
 - D I T I S * D K C I V C S N Y L W R K E
 - T S Q F P E T N V L S V V I I C G E K K
 4981 - AGTTCTCTGTGTTATAACCAAGAAGTGGCCATTAAACACAAAAACCCCTCACGGAGGA - 5040
 - S S S V * T K K C H * T Q K H L H E G
 - V P L C N K P R S A I K H K N T F T R E
 - F L C V I N Q E V P L N T K T P S R G K

FIG. 12 Con't

61/106

5041 - AGTATGCTTGCCTTCATGACAATTGCTGGCGCTGTTGAAGTTCCTCCTGGGATG - 5100
 - S M L C L H D K L L A L W * S S S P G M
 - V C F P A F M T N C W R C G E V P P L L G W
 - Y A L P S * Q I A G A V V K F L S W D G
 5101 - GCACATACGTGACATGTAGGAAGACRAACCACTGGGGCTGCTTGCGAAGGACATAA - 5160
 - A H T * H V G R Q H H A G L L V G R T *
 - H I R D M * E D N T M R G C L W E G H K
 - T Y V T C R K T T P C G R A C G K D I R
 5161 - GGTGGTAGCCCTTCACAAAAGTCACACTCTTTGATGTCACAGAACACACTCAGACA - 5220
 - G G S P F F H K S Q L S Q L F L I V Q E H T Q T
 - V V A L S T K V N S F * L S K N T L R H
 - N * P F Q K P S T L F D C P R T H S D I
 5221 - TTCTAGTAGCGAACAGATTAGCAGAACGCCCTGATTCAGCACGCCCTGATTAGTTGTG - 5280
 - F * * Q Q D * Q K P * F Q Q P * L V V V
 - F S S S K I S R S P D F S S P D * L L C
 - L V A A R L A E A L I S A A L I S C C V
 5281 - TTACATAGTTGAAAGCTTGAGTCTGCTGTAACTTACCTGTCAATTGTTACCTCCG - 5340
 - L H R F E G F E V C L * L T C Q F V P P
 - Y I G L K A L K S A C N * P V N L Y L R
 - T * V * R L * S L P V I N L S I C T S A
 5341 - CCTCGACTTATCAAGTCGCGAAAGGATATCATTAGCACACTGAAATTGCAACCAAAT - 5400
 - P R L Y Q V A K G Y H L A H L K L H Q N
 - L D F I K S R E R I S F S T L E I A P K L
 - S T L S S R E R I S F S T L E I A P K L
 5401 - TAGAGCTAAGTTGTTAACAGTGTGTTTAATGCTTGACATCTGGTTAACACAGCTCT - 5460
 - * S * V V * Q V C C L M L E H S G * Q R L
 - R A K L F N K C V * C L S I L V N N V L
 - E L S C L T S V F N A * A F W L T T T S C
 5461 - GCACCTGCCAAATGCGAGTGTGTTGTGAAAGTCTGAAATTGACTAATCGCT - 5520
 - R A C P M Q L M L L * V I L E F D * S P
 - Q L A Q Q C S * C C C K C * F L N L T N K L
 - S L P N A V D V V V S D S * I * D I A L
 5521 - CGTTAAATGGTGGCGATTGGTTGGTCTCATAGACAGAACATTGGTAACTCCGAA - 5580
 - C * I G W R F V F G S H R E H F G * L Q
 - V K L V G D L F V L L I E N I L G N S N
 - L N W N L A I C F W F S * R T F W V T P M
 5581 - TGCCATTGAACTTATGCGATTGGCATAGCAAAGGTATTGAAAGACGACGCCAGCRC - 5640
 - C H * T Y M P F A * Q K V F E E Q R Q H
 - A I E P I C H L S K R Y L K S S A S T
 - P L N L Y Q R W Q H C K S I E L L R T N E
 5641 - CAAATGCCCCATCCAGCAGTGGCAGTACCACTAACTAGACGAGCAGTGTAGGCAGCACATCA - 5700
 - Q M S I Q Q W Q Y H * L E Q Q C R Q Q S
 - K C P P S S S G S T T N * S S S V G S N H
 - N V H P A V A V P L T R A A V * R A I I
 5701 - TATCATCAGTAGGAGGCTGGCAGACTGTAACTGACATTGAACTTCGGCCACAAATGA - 5760
 - Y H Q * A E V A T L * V H * T S A H K *
 - I I S E Q R W Q H C K S I E L L R T N E
 - S S V S R G G N T V S P L N F C A Q M R
 5761 - GATCTCTAGCATTAATATCACCTAGGCATTGGCCATATTGCTTCATGAAGGCCAGCATAG - 5820
 - D L * H * Y H L G I R H I A S * S Q H Q
 - I S S I N I T * A F A I L L H E A S I S
 - S L A L I S P R H S P Y C F M K P A S A
 5821 - CGAGTGTCACTTATTAAGAGCAAGTCCTCAATAAAAGACCTCTTAGTTGGCTTAGAG - 5880
 - R V S P Y * R A S P Q * K T S * L A L E
 - E C H L I K E Q V L N K R P L S W L * R
 - S V T L L K S K S S I K D L L V G F R G

FIG. 12 Con't

62/106

5881 - GGTCAAGGTTAATTTGTAAGAAATTAAAACCACCAAAATTTCAAAGTTGGGGTTTGT - 5940
 - G Q V I F V K N * N H Q N I S K L G F C
 - V R * Y L * K I K T T E K I F Q S W G F V
 - S G N I C E K L K P P K Y F R V G V L Y
 5941 - ACATTTGTTGACTTCAGCCGACACTTCACCGTGTCTGCCATCCTGTTCAGCAGCAATAC - 6000
 - T F V * L E R T L H T V C C D P V Q Q Y
 - H L F D L S E H F T C V A I L F S S N T
 - I C L T * A N T S R V L R S C S C S A A I P
 6001 - CTGAGAGTGACGATTTAGTTGTCAGAAAGCTACCATATTGGAGAAGCAAATTAGCAC - 6050
 - L R V H D L V V C K S Y H I G E A N * H
 - * E C T I * L C A K A T I L E K Q I S T
 - E S A R F S C K L P Y W R S K L A H
 6051 - ATTCACTAGAATCTCCGAGATGTCATATTACATCTACGGAGGTTTAGCCATAGAAA - 6120
 - I Q * N L R R C T Y Y N L R R F * P * K
 - F S R I S A D V H I T I Y G G F S H R N
 - S V E S P Q M Y I L Q S T E V L A I E T
 6121 - CAGGCATCTTCTGACTATGCTAATGAAAGTTAGTAGGTATAGCATAGTGTTTAT - 6180
 - Q A L L L * C * L K S * V * V * Q W C
 - R H Y F C S N A N * K V S R Y S N G V I
 - G I T S V V M L I E K L V G I A M V L L
 6181 - TAGAGTARGCAATTGACTATCAGCACCTAAAGACATAGTATAGCCACATAGATTTT - 6240
 - * S K Q L N Y Q H L K T * Y K P O * I F
 - R V S N * T I S T * R H S I S H N R F L
 - E * A I E L S A P K D I V * A T I D F W
 6241 - GGCATGACTACTGATTAAGAACATGTAATGTAATGACATAGCACAAATGCCAGCTCAATAG - 6300
 - G * Y V I K K L Y G N * H K C Q L Q *
 - A S T T * * R N C M V T S T N A S S N R
 - L V L R N K E T V W * L A Q M P A P I G
 6301 - GAATGTCGCACTCATAGAGGTGTCAGCATGTCAGCTCTATAGACAGCCCTGCTGAG - 6360
 - E C R T H K K C R H A Q L L * D S L L E
 - N V A L I R S V D M L S S Y K T A C L S
 - M S H S * E V S T C S A P I R Q P A * V
 6361 - TCTGGAAATCATGTTCCAGATAGRATATATGGCCAAAGCTGGTGTGAGTTGATCTGCAT - 6420
 - S G I H C F Q * N I C A K L V * V D L H
 - L E Y I V U S S R I Y A P S W C E L I C M
 - W N T L F D P V E Y M R Q A G V S * S A *
 6421 - GAATTCGTAGAAACATCAGTCGCTTAACTCTGTATAGACAGCAACTTCAGATG - 6480
 - E L L * K H Q C S * H L D I E Q Q L Q M
 - N C C R N I S A V N I L I * N S N F R *
 - I A V E T S V Q L T S * Y R T A T S D E
 6481 - AACATTTGCTCCAGGATTAATTTACACTACCCCCAAAGAGCAAGGTGAATGCTA - 6540
 - K H L F Q V * L H L H P Q K S K V K C L
 - S I C S R C N Y T Y T P K R A R * N V *
 - A F V P G V I T L T P P K E Q G E M S N
 6541 - ATATTCAGATTTAGGACTCGAACGGAACTCAGTAATCAGAAACATCACGGCCAA - 6600
 - I F Q M F * D L E R N Q * N Q K H H G Q
 - Y F R C F R I S N G I S E I R N I T A K
 - I S D V L G S R T E S V K S E T S R P N
 6601 - ATTTGTTGAATGTTGAATTCCTTGTAGAGAAGGTAAACACCCAGTACAGTGACTC - 6660
 - I V E M V E I S L K K E L T H Q Y Q * V
 - L L K W L K S L * R R S * H T S T S E S
 - C * N G * N L F E E G V N T P V P V S P
 6661 - CATTAAATAAAATGACACRTGGTCTTAAAGGTCACTGGATATTGTCAC - 6720
 - H * N * N * H T G S * * G Q W I I L V R
 - I K I K I D T L V L N K V S G * F W S T
 - L K L K L T H W F L I R S V S D N F G P Q

FIG. 12 Con't

63/106

6721 - AACCGTGGCCGGTCATTTAAAAGTCRAAAGAAAGTACTACAACCTGTAAAGGTGGT - 6780
 - K P W P V H L K V Q K K V L Q L C K V G
 - N R G R C I * K F K R E Y Y N S V R L V
 - T V A G A F E K S S K E S T T T L * G W *
 6781 - AGCCAATGCCAGTAGTGGTGTAAAACCAATAATCATTAAATGCCAATTAACATTAAAGG - 6840
 - S Q C O * W C K N H N H L M A N N * E
 - A N A S S G V K T I I I * W P I T I K S
 - P M P V V V * K P * S F N G Q * Q L R A
 6841 - CAGGTGGGTCAGGTTTGCATCAGGGAGGAAGGCACATAGTATGTCTCTCAA - 6900
 - Q V G C K V C H Q G R K A H * I C L S Q
 - R W G A R F A I R G E R H I R Y V S L K
 - G G V Q G L P S G E K G T L D M S L S K
 6901 - AGGGCTTAACTGCTTCAAGTGTAAAGATACCTATTTATAATAATTACAGTTGAG - 6960
 - R A * A C H V * D T Y I Y N Y N Y Q L K
 - G P K L A M S K I P I I I I T S * S
 - G L S L P C L R Y L Y L * L * L P V E V
 6961 - TAGCATCATGTTCTCAGTAGTCTCAAGCAAGGCACRACCATGAAATCATCTGGCAATT - 7020
 - * H Q C S * Y S K Q G H N P * N H L A I
 - S I N V P S I P S K D T T H E I I W Q F
 - A S M F L L F V Q O A R T Q O P M K S S G N L
 7021 - TATAATTAAATCAGCAATACACCGNTTGTCTGGCTATTGTCTTACATCATCTC - 7080
 - Y N Y N Q Q * H Q F V L A I F V L H H L
 - I I I I S S N N T S L S W R Y L S Y I I S
 - * L * S A I T P V C P G A I C L T T S S P
 7081 - CCTTGACTACAAAAGAATCTCATAGACATTGGAGAACAGATCATCTAGTGG - 7140
 - P * Q K N L H R H W R S K D H S T * W
 - L D Y K R I C I D I G B A K I I I Q L S G
 - L T T K E S A * T L E K Q R S F N L V A
 7141 - CAGAACCCATAGRCCTAAAGGTGAAAAAATGTTGACTGTAGACACAGATHT - 7200
 - Q K R H S T * R L K K M L S C R A Q S N
 - R N P A I A L K G * K K C * V V E H R V I
 - E T P * H L K V E K N V E L * S T E * S
 7201 - CAGCACACAAATTAGAAATTTTTTCTCTCCCATCATAGACAGAGGGAAATTAGTAG - 7260
 - Q Q H N * K F F S P M N H R Q K G I * *
 - S N T I R N F F S L P C I D R R E F S S
 - A T Q L E I F F L S H A * T E G N L V A
 7261 - CATTAAAAACCTCTCAAAGGACACAAAGTTGTATATTAGGGAAATCTCACACATCTC - 7320
 - H * K P L Q K D T S L * Y * G I S S Q H L
 - I K N L S K R T Q V C N I R E S H N I S
 - L K T S P K G H K F V I L G N L T T T S P
 7321 - CTGAGGGGAAACCACTGAAATCTGAGGTCTGTAATTCTCTTGTCAATCTCARAGCT - 7380
 - L R E Q P * N * R S G K F L C Q S Q S S
 - * G N N T P B I R G L V N S F V N L K A L
 - E G T T L K L E V W * I P L S I S K L L
 7381 - TRACAGGACATTGGTGTCAAGCAAGTGGATTITGGAGAACATCACACRGATCTGTGATIG - 7440
 - * Q S I * V Q V D F E N N Q Q H L * L
 - N R A F E F S K W I L R T I N S I C D C
 - T E H L L S S A S G F * E Q S T A S V I V
 7441 - TACCAATTTCATCATCTGAGCATAAATGTTGAGCTTAAATAGCCACAAATAGG - 7500
 - Y H F H T * A * M * L A L N S O Q N R
 - T I F I I L E H K C S W L * I A N K I G
 - P F S S Y L S I N V V G F K * P T K * A
 7501 - CTGCAAGCTGAGCTGGCCCAATGCTTGTAGCAGGTAAAGGCTGTAAAGATGGCTCAA - 7560
 - L Q L T C P F K C L E Q V K R L * E W L *
 - C S * R A P N V L S R * K G C K N G S K
 - A A D V P Q M S * A G E K A V R M A L K

FIG. 12 Con't

64/106

7561 - AATTTGTAATGTTAATCCAAGAGGCACCTTAAATAGGTTCAAGTGTAAACAG - 7620
 - N L * C * Y Q E A T * K * V S K C * N Q
 - I C N V N T K R O L K N R F Q S V K T R
 - F V M L I P R G N L K I G F K V L K P E
 7621 - AAGGTAGATCAGGAACACTACATCTATAGGTTGATAGGCCCTATAAACATAGAGAAACCAT - 7680
 - K V D H E L H L * V D S P Y K H R E T H
 - R * I T N Y I Y R L I A L I N I E K P I
 - G R S R T T S I G * * P L * T * R N P S
 7681 - CTTTTTTTTAACACAAACTCTGTAAGTTAAATTACTGACTTTCTGAAACAT - 7740
 - L Y F * T Q T L V S V * N Y L T F L K H
 - F I F K H K L S * V F K I T * L F * N I
 - L F L N T N S R K C L K L P D F S E T S
 7741 - CAAGCGAAAAGGCATCAGATATGTACTCGAAGTCRATTAAATGCATTATCGAATATCA - 7800
 - Q A K R H Q I C T R K C N * M H Y R I S
 - K R K G I R Y V L E S A I K C I I E Y H
 - S E K A S D M Y M S K V Q L N A L S N I I
 7801 - TAGTATGTCGTCGTCAGGGTTAGAACAGCACAAAGGTTGTCACACAATT - 7860
 - * Y V S W V * K Q R K G C H T I
 - S M C L C T H G F R N S K E R V V T Q F
 - V C V C V P N G L E T A K K G L S H N S
 7861 - CAAAGTTACATGCTGSTATARACACRTTAGAAGATGTTAATATAATCAGCGACTTG - 7920
 - Q S Y M L V * Q H * * N C * * * S P T V
 - K V T C S Y N N I S R I V N N N H R L *
 - K L D H A R I T T L V E L L I I I I T D C D
 7921 - ACTTGTGTTGTCATGGPAGAACCCAAAAACCCACACCGGACACATTGATTCTCTGIGG - 7980
 - T C C S W * N Q K P N H G Q H L I S L W
 - L V V H G R T K N P T T D N I * F L C G
 - L L F M V E P K T Q P R T T F D F S V A
 7981 - CAGCAAATAAACCATCTTAAAGGTATGACAGGGTTGCCRAACGTATGATTAAATAG - 8040
 - Q Q N K Y P * K V * Q G C Q T Y D * *
 - S K I N T I L K R Y D R V A K R M I N S
 - A K * I P S L K G M T G L P N V * L I V
 8041 - TATGAAACCCGTAACTTAAAGGAAATAATGGAGAAATAATCCTGAGTTAAATAAGAG - 8100
 - Y E T L * H * N K M E E I N P E L N K E
 - M K P C N I R I K W K * I L S * I K S
 - * N P V T L E * N G R N K S * V K * R V
 8101 - TGTCGTATCATAAAATTCTACAGGATAGTAACACCCCTCATAGATGAGTATGTTGAG - 8160
 - C L I * K F H Q D S K F P P S * M K Y V E
 - V * S K N F I R I V N P P P H R * S L S
 - S D L K I S S G * T P L I D E V C C * V
 8161 - TGTAATTGGAGCTAACATCATCAAAGTGGTGACCGGTCAAAGTCACTACCACTAG - 8220
 - C N * E L E H H Q K N C T G Q G H Y H *
 - V I R S L N I I K S G A P V K V T T T S
 - * L G A * T S S K V V H R S R S L P L V
 8221 - TGAGAGTAAGAAATATAGGAAATAAACATGTTCTTTAGTGTAAACAGAAATAC - 8280
 - * E * E I I K * T C S F S C * Q E Y H
 - E S K K * * E N N H V R L V V N K N I T
 - R V R N N K K I N M F V * L L T R I S L
 8281 - TTGAACCCACACTCTGTTCTCTAAATGATAAGCTTACCTTTTCAGAAGAGAAAT - 8340
 - L K P Q L C C F L * * * A Y L F P E E N
 - * N H N S V V F S N D K P T F F Q K R I
 - E T T T L L F S L M I S L D P F F S R R E *
 8341 - AATTCATATCATGATTGATTCCTCTAAAGAGACATTACAGCGTTCTTAAATTAA - 8400
 - K S Y H * F D S P * E T L Q Q F L L I *
 - N H I I D L I L L K R H Y S S S S * F K
 - I I S L I * F S L R D I T A V P L N L R

FIG. 12 Con't

65/106

8401 - GAGGAATTTGCTCATGICAAGAGTGAATAGGAAGAACACTGGATAGGATTGTGTTCC - 8460
 - E E I C S C Q R V N R K T T G * D L C S
 - R K F A H V K E * I G R Q L D R I C V P
 - G N L L M S K S E * E D N W I G F V F L
 8461 - TCCAGAAAATGATGTTAGCATGCATGGTATGCCATCAATTGTTCTTCGGCTGCCAA - 8520
 - S R K C S * H A W Y S H Q F V P S A C Q
 - P E N V V S M H G I A I N L F L R L A K
 - Q K M * L A C M V * P S I C S F G L P R
 8521 - GATAGTTAGCCCCAATRAAATGCTTCGATGATGATGATTTACATTGTAACAAAG - 8580
 - D S * P Q L K M L P M M M H L H L * Q K
 - I V S P N * K C F R * C I Y I C N K S
 - * L A P I K N A S D D D A F T F V T K A
 8582 - CTGTCACCATGAGAATGCCATAGCTTAAAGGTAGCATCCAGATGCTG - 8640
 - L S T M R N G P * A C K G Q H S K N A L
 - C P M P * E M A H K L V K V S I P R M L C
 - V H H E K W P I S L * R S A F Q E C S V
 8641 - TTATCTTACAGCTATAGAACCCCCAGGCTAGTTTTGCTTTAAATCCACAGAT - 8700
 - L S L Q * N H P G L V F A L * I H T D
 - Y L L Y S Y R T T Q O G * F L L Y K S T Q I
 - I F T T A I E P P R A S F C F I N P H R *
 8701 - AAGTGAACAAACCCCTCTTAGAGTCATTCCTTTGTCACTGTTGGCTCTAGGTICAT - 8760
 - K * K T L L * S H S L L S H V W S * G H
 - S E K P F F R V I L F C H M F G P R V I
 - V K N P S F V T C L V D G S Y
 8761 - ACATATCGCTAAATAAGTCCTATTATTAGCCCTATGATCTGTCACAGTCCTCA - 8820
 - T Y R * * * G P I Y * P Y V L L H S L Q
 - H I A N N K V P F I S R M Y C C T V S N
 - I S L I I R S H L L A V C T V A Q S P I
 8821 - TTAAAGTGAATCTCGCTGGAGACGAGTCATTAARGATCTGAATCGRCAAGTAGTGTC - 8880
 - L K * N L R R R S H * D L N R Q V C
 - * S R I C V G D E V I K I * I D K * C A
 - K V E S A S E T K S L R S E S T S S V P
 8881 - CAGTTGGCACCCATGCTGGACACAGCTGTACCTGGTGCACACTCCCTTATCAGAGCCG - 8940
 - Q L A T I V * A Q L Y L V Q L L Y Q S Q
 - S W Q P L S E H S C T W C N S F I R A S
 - V G N H C L S T A V P G A T P L S E P A
 8941 - CACCAAGTGAATACTCTCATGTGTAGGGTACAGCTAAAGTAAGTGTATTAGTATT - 9000
 - R Q S E * L S C C C V Q L K * V Y L S I
 - T K V N N S H V V G Y S * S K C I * V L
 - P K * I T L M L * G T A K V S V F K Y *
 9001 - GACACAGTTGAGTATCTTGGACATTCATCATATTCCCTTTGGTAAACAGCATTT - 9060
 - D T V E Y T L R H S S L F L L V * Q H F
 - T Q L S I L C D I H H Y S F N W Y N S I F
 - H S * V Y F A T F I I I P F G I T A F S
 9061 - CACCATATTCTGAAGGTACACTTTCAGAACATCTTGGTACAGCTG - 9120
 - H H N S E G R T F Q E A F F A S C T S *
 - T I I L K V T L F K K H S L H L V Q V R
 - P * F * R S H F S R S I L C I L Y K L G
 9121 - GCATCGCACACCTGGTGCACGCTTGACTTGCTGTAGTTGGTAGAAAGTTCAA - 9180
 - A S Q H L V A T L D L L V V L G R R F Q
 - H R N T W L P R L T C L * F W V E G F N
 - I A T P G C H A * L A C S F G * K V S T
 9181 - CATGTCACCATGTTACACCAAGCATGAATGAAATTCAGCATAGTCATGTAACCTG - 9240
 - H V H P Y T K A * M K F Q H S Q L * P *
 - M S I L T P K H E * N F S I V N C N L D
 - C P S L H Q S M N E I S A * S I V T L T

FIG. 12 Con't

66/106

9241 - CCACTTTGAAATCACTGACAAATCTGTGACTTATTATCTGCACAAAGTCATCAAGTA - 9300
 - F L L K S L T N L V T L L S R Q S H Q V
 - H F * N H * Q I L * L Y Y L D K V I K *
 - T F E I T D K S C D F I I S T K S S S K
 9301 - RAAGATCAATCACAGAACACACRCACTTGTGATGAAACCTGTTGGCATCTGTATGAAGT - 9360
 - K D Q S Q N T H I L M N L F A H L L + S
 - K I N H R T H T F * T C L R I C Y E V
 - R S I T E H T H F D E P V C A S V M K *
 9361 - AATTTTCACTGTGCTGTCATAGGGATAAAATCCTCTAATTAACTGGTGAATCTGTG - 9420
 - N F S L C C P * G * N P L I * V V N L V
 - I F H C A V H R D K I L * F K W * I L *
 - F F T V L S I G I K S S N L S G E S C E
 9421 - AGCGCTTGGCTARGCCATCATTAATGAAAGRCGCCAAGTTGTCATGCTGAAATCTC - 9480
 - S A W L S L S L N E D R Q V V H D = N L
 - A L G * A Y H * M K T A K L S M T E I S
 - R L A K P I I K * R P P S C P * L K S P
 9481 - CATAAACGATGGTTCGRRGCATAGCCTCGAGCTTATCGCTGTATGAAATCATCCA - 9540
 - H K R C V V R H S P R A Y I A V * I H P
 - I N D V F E G I A L E L I S L Y E F I H
 - * T M C S K A * P S S L Y R C M N S S I
 9541 - TAGCGAGCTCGAGAARGTCAGTTCCATTGTGATCTGGCTTAAATCCTCTAAGTCTC - 9600
 - * R A R E S Q F P F V I W A * N P L S L
 - S E L E K V S F H L * S G L K I L * V S
 - R S R K S V S I C D L G L K S S K S L
 9601 - TGCTCTGTAAGTAGGTTCTGGCAACTTGATTAATGCGCTCTTCTTCTAAGT - 9660
 - C S E * S R F Q A T V E * C R L L S * S
 - A L S K V G F R Q L L N N A V Y F L K V
 - L * V K * V S G N C * I M P S T F L K *
 9661 - AGTTARACTGTTTTACTGATCTCCAATTAAATGACTCCRTTGACGCTGAGCTTGTG - 9720
 - S * T V F P L L I L Q L M * L H * R * L V
 - V K L C F Y * F S N * C D S I D A S L C
 - L N C V F T D S P I N V T P L T L A C A
 9721 - CTGGTCTCTTGTAGGTTCTGGACTCTTGTGACTGACCTCTGTATATAAAACACCAATTAC - 9780
 - L V P L K V L D L * L N L L L K H Y
 - W S L * R C T F D * T F C Y * N T I T
 - G P F E G V R P L T E P S V I K T P L R
 9781 - GGGGTTCTAAAAGGTCTACCTGCTCTTCACTCTACCATCAAACAGCAGTAAGTG - 9840
 - G R F * K G L P V L P L Y H Q T R Q * V
 - G V S K K V Y L S F H S T I K Q D S K *
 - A F L K R S T C P S T L P S N K T V S E
 9841 - AAGAACAGCACTCTAGTAGGTTCTGGCAATGTGACTCATGGCACAGCACCTATIG - 9900
 - K N K H S Q * V S W O C Q S L C R H L L
 - R T S T L S R F L G N V S H C A D T Y C
 - E Q R S V G F L A M S V I V Q T P I V
 9901 - TAGATACATGTGCTGGGGCTCTCTTTGTAGTCCAGATTAACAGTATTAGCAGCATAT - 9960
 - * I H V L G L L F C S P R L Q Y * Q R Y
 - R Y M C W G F S F V V P D Y S I S S D I
 - D T C A G A S L L * S Q I T V L A A I S
 9961 - CAACACCCAAATTAGTGAATCTTATCTCTGGCACTGGTTAAATGTTAGGCTTAGCCC - 10020
 - Q H P N Y * V S * S L A L V * C Y A * P
 - N T Q I I E Y L N L W H W F N V T L S P
 - T P K L L S I D I S G T G L M L R L A Q
 - AAAGCTCAATGCAACATTAACAGGAGTGTGCTTATTTCAAGATCTCACATCAA - 10080
 - K A Q M Q H * Q E V L S Y F Q R S P H Q
 - K L K C N I N R E K C C L I F K D L H I N
 - S S N A T L T G S V V L F S K I S T S I

FIG. 12 Con't

67/106

10081 - TACCATCTACCCTTGATGAAACAGCATTATAATGATGGAAACAGGTGTTCCGGCGGT - 10140
 - Y H L P L C K Q H Y * * W K Q V L R R R
 - T I Y L C V N S I I N D G N R C F A G V
 - P S T F V * T A L L M M E T G A S P A C
 10141 - GTCCCAAAAGTGTCTTTAACAACATTATAGCCACATTTCATAACTCTGTAAACC - 10200
 - V H Q S V L Y * Q H Y K P H F L N S V T
 - S I K V S F I N N I I S H I F * T L * P
 - P S K C P L I T T L * A T F S K L C N L
 10201 - TGGTAATGTTCCACAGGTATAGTATCAAATTGTTGAATCCATAGGCTAAAC - 10260
 - W M Y S T G Y K Y Q I V C K S I G * I
 - G K C I P Q V I S I K L F V N P * A K S
 - V N V F H R L * V S N C L * I H R L N P
 10261 - CACCAAAATCATCATATAATAGCATCCAGTACTGTGGTACTCATGGTGTCTC - 10320
 - Q Q K S S Y Y M H P S T V G T H L H G V
 - S R N H H I I C I Q V L S V L I C M V S
 - A E B I I L Y A S K Y C R Y S F A W C L
 10321 - TGCAACGACCAACCTAAATTCGATCGTAAATACAGTAGCAGATTGAGTGGAACT - 10380
 - C K Q H R L N C I V * Y T * Q I * V E H
 - A N S T T * I A S C N T R S R F E W N I
 - Q T A P P K L H R V I H V A D L S G T *
 10381 - ARTCAATATCCGACACTACTTGTGTTGCTAGAAGCTCACRAGGACTATCAGAATAGTAA - 10440
 - N Q Y F T L L V C H E T H K D Y Q N S K
 - I N I R H Y L F A M R L T R T I R I V K
 - S I S T C L P * D S Q G L S E * * K
 10441 - AGAAAGGCCATTGCTTAAATAGTAAATGCACCTTATCGAAAGCTGGAGTGTGGATG - 10500
 - R K K A I A L N * * M H F Y R K L E C G M
 - E R Q L L * I S K C T F I E S W S V E C
 - K G N C F K L V N A L L S K A G V W N A
 10501 - CATGCTTATCACATCACAACTACACCATCACAGCGCTGGTAAGTTCAGAAGTGGAAAGA - 10560
 - H A Y S H T N Y H H H S L V S S S L T R
 - M L I H I Q T T T I T A W * V Q V * Q D
 - C L F T Y K L P P S Q P G K F K F D K T
 10561 - CTCCTGTCACACCTACACRATGTCATGGCTGGTAACGATCAGCTTACATTCC - 10620
 - L L C Q T Y T Q L H W G N D Q R Y N S
 - S C V K F T H N C I G W V T I N V T I P
 - L V S N L H T I A L A G * R S T L Q F Q
 10621 - AACACACACACACCATCACTGAATTATGATGTTGAGCATAAAGAATAGAACATT - 10680
 - K T N K H H Q * I Y R D V * R K N R R V
 - K Q T N T I S E F I V M C S I R I E E F
 - N K Q T P S V N L S * C V A * E * K S S
 10681 - CCTCTATTGTTGAGCTTGTCACTACATGGCTGGCATCGTAGAACCTCCATTCTACT - 10740
 - P L F C K L C H C Y M A E H R R T S I L L
 - L Y F V S F V . T T W L S I V E L P F Y F
 - S I L * A L S L H G * A S * N F H S T S
 10741 - CAGCTGAGGACACACTGTAGACCTTGGATTCTCATGAGAACCTGAACT - 10800
 - Q P E A H T * * P L D F Q C H E E L E T
 - S L R H T L D S L W I S N V M K N W K L
 - A * G T H L I A F G F P M S * R T G N L
 10801 - TATCAGCAACGAACTGCAAGACTCACRACCATGTTGACTTTCTGGAAAGCAGAAATTAA - 10860
 - Y Q Q A M Q T S Q P C V V L F C K Q N *
 - I S K Q C R L H N H V L Y F S A S R I N
 - S A S M A D F T T M C C T F L Q A E L T
 10861 - CCCTCAGTCATCTCTATAATAGGTATTCAACAGACCATCACGGCTTAACAAAGC - 10920
 - F S V H L L * G I Q Q T N Q R A * Q S
 - P Q F I S Y N R V F N R P I N A L N K A
 - L S S S P I G Y S T D Q S T R L T K H

FIG. 12 Con't

68/106

10921 - ACTCATGGACTGCTAAACATCTAGTCATGATAGCATCACAACTAGCCACAATGTGCATTIC - 10980
 - T H G L L N I * S * * H H N * P R V H F
 - L M D C * T S S H D S I T T S H M C I S
 - S W T A K H L V M I A S Q L A T C A F P
 10981 - CATGTCACCTGGCAATGTGGTCATGGTACTCTGAAGGTACCGCTAACGCCCACTGCT - 11040
 - H V P G N V G H G Y S E G Y P * S P T A
 - C T W Q C W S W L L * R L P V K P H C *
 11041 - GARACATCACTATAAATGGGTTATAGACATAGTCRAAACCCACAGRATGTTCCACGAGG - 11100
 - E H Q S * M G Y R H S Q N P Q N D S S R
 - N I N H K W V I D I V K T H R M I P A G
 - T S I I N G L * T * S K P T E * F O Q A
 11101 - CATAAGTATCTGTAGAAGTAGAAGCAGTGGCACGCTTGTACACAGCAACACGTTIC - 11160
 - H K Y L H M K * K S K L H V C H T D N T F
 - I S I * * S R K A S C T F V T Q T T R S
 - V S D E V E S K Q V A R L S H R Q R H V L
 11161 - TTTCAGGTCCAATCTGTGCAAGTACTTCATTGTTAGCTCAAAGCCATGCCCCCCAA - 11220
 - F Q V * S Q S T S L M * A Q S H A P K
 - F R S N L D K V L H * C K L K A M R R P K
 - S G P I L T K Y F I D V S S K P C A Q R
 11221 - GGACGACACGACTCTGTCTGCAATCCTTCTAGTGTACTGACGCAATTGTTACTATCT - 11280
 - G R T R L C L T I L S V Y H * A F V L S
 - D E H D S V * Q S F Q C I T E H L Y Y L
 - T T N T L S D N P V S V S L S I C T I L
 11281 - TAATACCGCAACTACATTCAGGGCAGGGCTTATACATGAGTGTATAAGCTTAAACT - 11340
 - * Y A L H S R A S L Y T * V V * D V * T
 - N T H Y I P G Q A F I H E W Y K M F K L
 - I R T T F Q G K P L Y M S G I R C L N W
 11341 - GGTCACTCTGGGGTTCTGCTGCAATCCTTCTAGTGTACTGACGCAATTGTTACTATCT - 11400
 - C H L V E V E D H * L W * I L C Y F Q C Q
 - V T W W R F C I T I S G E F C V I F F S V N
 - S P G G G F A L T L V N S V L F S V S T
 11401 - CATAACAGCTGGTAACAGCTACTAGTACACTGTRAGAAAATCTCTACTGAGAGCTA - 11460
 - H N Q S V Q L L S * H L * K I L A G E V
 - I T S R Y S Y * V N T T C R K S * L E R *
 - * P V G T R T K L T P V E N P S W R G R
 11461 - GGTTRGTACCCAGCATCTCTAGTTGATGACAGCCCTCATACATCAAAGCCCATCCAG - 11520
 - G * Y P Q H D * L H U S P L H Q S S Q S T
 - V S T H S I S S C M T A L Y I K A N P R A
 - L V P T A S L V A * Q P S T S K P I H A
 11521 - CACCAAGCTGGACAGAATGCTCTTCCGGGGTGATTAACATATTAGGGTACCCATTGACT - 11580
 - H E R D E * L L R G * T Y * G N H * L
 - T N V T N S F F A G D K H I R V T I D L
 - R T * T R I A S S R V I N I L G * P L T W
 11581 - GGTAAATTCTATTGAAACCCATCATAGAGATGACTCTACGGTAGGTCAATGCTTGGTA - 11640
 - G N S F * N P S * R * V Y G R S C P L V
 - V I H F E T H H R D E S T V G H V L W Y
 - * F I L K P I I E M S L R * V M S F G M
 11641 - TGCCCTGGATGTGCAACACATTAATCCTTCACTGTTGAATTTTATATCAACGGCTGAGGTG - 11700
 - C L V C Q H I I L Q S * I L Y O R * G V
 - A W Y V N T * S F S L E F Y I N A E V C
 - P G M S T H N P S V L N F I S T L R C V
 11701 - TAGGTGGCTGGTAGGATGAAGACCACTAATGACTCTACTACAGTCTTAAAGTCNCG - 11760
 - * V P V * D E D Q * * S Y Y S P * K V Q
 - R C L C R M K T S N D L T T V L K K S S
 - G A C V G * R P V M I L L L Q S L K S P V

FIG. 12 Con't

69/106

11761 - TTACATTTCTGCTTGTATGTAGCCACATTGCCACGTGGTATTTCTAGACTTGTAATT - 11820
 - L H F L L V M * P H C D V V F L D L * I
 - Y I F C L * C S H I A T W Y F * T C K L
 - T F S A C N V A T L R R G I S R L V N C
 11821 - GCAGTTGTCATAAAGATCTCATCAGACATTATGCCAAATGCCATTGGCC - 11880
 - A V C H K D L Y Q T L C T K C Q F L P L
 - Q F V I K I S I R H Y A Q N A N F C P C
 - S L S * R S L S D I M H K M P I F A L V
 11881 - TGATAGGCCACATGAAGCGGTGACATACRAGATGTCCTTCAGTAGTTGTC - 11940
 - * * P H * S G H Y K S V L F Q * F V *
 - D S H I E A V D I T R V C C F S S L C E
 - I A T L K R L T L Q E C A V S V V C V N
 11941 - ATATGACATAGTCATACTGCACCCCTGTGATGAACTAACAGTCGCGTAGGCATCCTA - 12000
 - I * H S H I Q N P V M H N Q Q S A * A I L
 - Y D I V I F R T L * I N S L R R Q S *
 - M T * S Y S E P C D E S T V C V G N P K
 12001 - AGATTTTGAAAGCTACAGCGTTCTGTGAAATTAAAGGTGAGATAAAAACAGCTTCTCC - 12060
 - R F L K L Q R S V N Y K V R * K Q L F S
 - D F E * S Y S V L * I I R * D K N S F S P
 - I F E A T A F C E L * G E I K T A F L Q
 12061 - AACGAGGATTGCGTGAAGAAATTCTCTACACGCCATTGGCTCTGGTCGATGAGCAG - 12120
 - K Q D V C * E I L L Q R L F E V C * L Q
 - S. R I A C K K F S Y N A Y L R S V D C R
 - A G L R V R N S L T T P I * G L L I A D
 12121 - ATGAAACATCATGTTAAACACCCCTTGACAGATTGGACATGACTAT - 12180
 - M K H H V * H L C R T F * S I Z L T Y
 - * N I M C N N T F V E H F E A L S * L I
 - E T S C V I T P L * N I L K H * A D L S
 12181 - CCTTGTGTCCTATGGCTATGCTACAATACAGACTCACAGTCAACATTTCAG - 12240
 - P C V U L L A Y C H K L K H S Q C O Q F Q
 - L V C F * L I V I N * S T H S V N N F S
 - L C A F S S L S * T K A L T V S T I S A
 12241 - CAGGACAACGGGGCACAGTCCTAACAGATGTCGACCTTATGTTCTAAAGTCTCC - 12300
 - Q D N G D K F Q G T C L D L L F S * V C
 - R T A T T S S K E H V W T Y C F H K S A
 - G Q R Q V P R N M S G P I V F I S L H
 12301 - ACATCTGAATAATTCGGTTCTAGTGCCCTTACTAGCAARTGTCGGGGGGCTG - 12360
 - T L N * N I L V L V U C L * S A M C G G G L
 - H * I K I F W F * C A F F S Q O C A G G W
 - T E L K Y S G S S V P L V S N V R G A G
 12361 - GTAAATTGACGGATCCCAATATAGACAGTGTGTTTCACGRAGTCTAGCATTGACAA - 12420
 - V I E Q D R Q Y R R S V L H E V * H * Q
 - * L S R I A N I D V V F C T K S S I D N
 - N * A G S P I * T * C F A R S L A L T T
 12421 - CACTCRACTCATATTAGTAGCCATAGAGATTCTACAAAGACTACAACTGTCAGCAGTTG - 12480
 - H S S H N * P * R F H Q R L Q C Q Q L
 - T Q V I I S S H R D F I K D Y N V S S C
 - L K S * L V A I E I S S K T T H M S A V V
 12481 - TTTCTGGCAATGCTTACAGTCGAGAAAACATACTGTGTTGAAATTCACTTGA - 12540
 - F L A M H L Q C R K H T V L V L N S L *
 - F W Q C I Y S A E N I L F * C * I H F E
 - S G N A F T V Q K T Y C S S V E F T L N
 12541 - ATTATCACAAACACTCTACGGCCGACAGGGATGATTTCTACATTCATTCATGG - 12600
 - I Y Q N T L R A H A Q V * F Y Y I Y L W
 - F I K T L Y A R T R R Y D S T T F I Y G
 - L S K H S T R A R A G M I L L H L S M G

FIG. 12 Con't

70/106

12601 - GCGAAATTTAATGCCCTTACATAGGGCATAACAGCTGCATGAGAGCATGCCGTAT - 12660
 - A N I L M P F H I G H Q O L H E S M P Y
 - Q I F * C L F T * G I N S C M R A C R I
 - K Y F N A F S H R A S T A R * E H A V Y
 12561 - ACACATTCGGAGCAGATGGTAATAGAGGACGARGCCGATGGCAAATGACTCTTACAC - 12720
 - T L C E Q M G N R E Q V R W Q N D S Y Q
 - H Y A S R W V I E S K S D G K M T L T S
 - T M R A D G * R A S P M A K * L L P V
 12721 - TACCGAGTGGCTTGAGTAGACTTGTGATGCGACCTTTGATAATTGCAA - 12780
 - Y Q V V L G G V * S T F P A C R C P F D N L Q
 - T R W S L E C R V L L H A D L L I I C N
 - P G G P W S V E Y F C M P T F * * F A T
 12781 - CATTGCTAGAAAACCTCTGAGATGTTGAGTGGCACAGGCACTTCTCACAT - 12840
 - C H C * K T H L R C * V L G T S Q * F S H
 - I A R K L I * D V E L C W V Q A S N S H I
 - L L E N S S E M L S V G Y K P V I L T *
 12841 - AGTGCTCTGGCACTAGACTAGAGTAGGYGCACTAACAGTGTGAGATGTCACA - 12900
 - S A S L V A L E * V H * V A L Q C E M S T
 - V L L W H * S R C T K W H Y S V R C Q H
 - C S C G T R V G A L G S G I T V * D V N T
 12901 - CAAAGTAACTACCAACATCCTACTGTATGTCGAGTACCTCTGTACACACAGCATC - 12950
 - Q S N H Q K H S T C M S * Y L C T Q Q H H
 - K V I T N I Q L V C R S T S V H N S I T
 - K * S P T F N L Y V V V P L Y T T A S P
 12961 - CATAGTCACCTTTCTRAAGGTGACTCTCACATGTACTTTACTATTITAGTACAC - 13020
 - H S H F Q R C T L Q S V L Y Y F * L H
 - I V T F F P K G V L D S N L Y F T I F S Y T
 - * S P F S K V Y S P I C T L L F D V T R
 13021 - GGTAACCTAGTAAAGACATGTTCTGTCATGTGGCTAGGTTTCCCACCTCCATG - 13080
 - G N Q * R H S F C S M V * V F Q P F M
 - V T S K D I V S V Q W N W S R F S N L P *
 - * P V K T * F L F N G G L G F P T S H E
 13081 - AAAGATGCAATTCTCTGTCAAGAGACTTCGCGTACAGTGGCAATACCATATGCAAGCT - 13140
 - K D A I L C Q R V L R V Q W Q Y H M T A
 - K M Q F S V R E Y F A Y S G N T I * Q L
 - R C N S L E S T S T R T V A I P Y D S L
 13141 - TAAATGTTCTCAGYGGCTTGAAGGCTTCTGTCGAAAAGCTGAGTCTCTAGT - 13200
 - * M F P Q W L * A F L L R K A * V V S Q Y
 - K C F L S G F E R F C C E K L E S L S T
 - N V S S V A L S V U S A A K S L S L S V Q
 13201 - AAGTGTGGCAAGTATGTAATGCGACGATTAGCTCAACATGTGCTATGGCATGTA - 13260
 - K C W Q V C N R Q H * S N H M D L L S H *
 - S V G K Y V I A S I S P I T C C Y R I E
 - V L A S M * S P A L V Q O S H V A I A L K
 13261 - AGTCAGTGAACATTGCACTGGCTACACATGTTTGTATAAACCAAAACCTGACCAT - 13320
 - S Q * H C H C L H M C F M C I N O K P D H
 - V S D I V T A Y T C V F V * T K N L T I
 - S V T L S L P T H V F L Y K P K T * P L
 13321 - TAGCCACATAATGGAACATAATGGAGGCTTATGACTGCAATATAAGCTCATACCTC - 13380
 - * H I M E N * W E A Y V T C N N S S Y L
 - S T * W K T N G R L M * L A I I A H T S
 - A H N G K L M G G L C D L Q * * L I P P
 13381 - CTAGATACAGTGTGTCACATGACATCACACCTGGGCAATTGCAACATAGGGAT - 13440
 - L D T V V S H Q * H H N L G H C K H R D
 - * I Q L C H I S D I T T W G I A N I G I
 - R Y S C V T S V T S Q P G A L Q T * G L

FIG. 12 Con't

71/106

13441 - TAACAGAACAACTAATTGTGATGTTGAATTGACATGGTCATAGCAGCACTTGCAC - 13500
 - * Q T L I C V M K * H G H S S T C N
 - N R Q H * F V * C * N D M V I A A L A T
 - T D N T N L C D V E M T W S * Q H L Q H
 13501 - ATAGGAATGGTCCTCTAACAGGCACCGCACCGTAACTGAGTCGTAATGCACTA - 13560
 - I G M V S * Y R H R N E V K S V N C T I
 - * B W S P N T G T A T K * S L * I A Q Y
 - R N G L L I Q A P Q R S E V C E L H N T
 13561 - CACAGAACACTAACGGCTCGAACAGCTGTATCTGGCTGTACATAGCCTCATAAACTCAG - 13620
 - H K H L Q P A R R L Y V V C T * P H K T Q
 - T S T Y S L Q D C M W C V H S L I K L R
 - Q A P T A C K T V C G V Y I R S * N S G
 13621 - GTTCCCAGTACCGTGAGGTCTATAGTAGCTTACCGAATACATGTCCACATG - 13680
 - V P S T V C Y H * D A L R N T C P T C
 - F V P V P * G V I I S * H Y G I H V Q H V
 - S Q Y R E V L S L V S I T E Y M S N M W
 13681 - GGCCAGTACGTCATCATGTAACCTCTAAAGTATTCGTAATACAAGTGAAGACATCAG - 13740
 - G Q * A H H V T F * C I V N T S E R H Q
 - A S K L I M * L S N V L * I Q V K D I S
 - P V S S C N F L M Y C K Y K * K T S A
 13741 - CATACTCTGATAGGATGTTCTGTAAGGGTAAGCATCAATAGCCTGACAGGAC - 13800
 - H T P D * D V L * V G K H Q * P V T R T
 - I L L I R M F C K W V S I N S Q * H E P
 - Y S * L G C F V S G * A S I A S D T N L
 13801 - TTTCATCATAGTGTACCATCTGTTTGACAAATCATGCCAAACAGCTGCCCTA - 13860
 - F Q S * V Y H F * O Y H R Q N S L R L
 - F N H K C T I C F D N I I D K T A C A *
 - S I I S V P S V L T I S S T K Q P A P N
 13861 - ATATTCCTGATGAGTCTGGTAGGCTAACGGCAGTACGTAATCATCTCTGTTTACAGCA - 13920
 - I F L M D L G K A G T R N H L L V * L A
 - Y S * W I W V R Q V H V I I S L F N * H
 - I L D G S G * G R Y T * S S P C L T S I
 13921 - TTGTATGCTGTGAGCRAAATTCTGAGGCTCTTACTAAGGTAGTCGTCAGTCCAACAT - 13980
 - L Y A T V S K I R E V L * * G Q S Q S N I
 - C M L * A K F V R S F S K V S L S P T F
 - V C C E Q N S * G P L V L R S V S V Q H F
 13981 - TTGCCATGAGACATGAAACATATTGATATAAGACTGCCTTAAGTCTTATAGT - 14040
 - L P Q T * T H Y F D N K E L P * S S * C
 - C L R H E H I I L I I K N C L K V L N A
 - A S D M M T L P * * * R T A L K F L M L
 14041 - TAGCTACTAACCTTGAGCCCATAGTTACTGTATAGCACACRACGGCATCATGAAA - 14100
 - * L L N L E P H S Y C Y S T Q R H H Q K
 - S Y * T L S R I V T V I A H N G I I R K
 - A T K P * A A * L L L * H T T A S S E R
 14101 - GAATCATCATGGAGAATGTTTACGAGGTAAAGCTAAACTCATCCACGAACTCATGAT - 14160
 - E S S W R N V Y A G K R K T H P R I H D
 - N H H G E M F T Q V S V K L I H E F M I
 - I I M E K C L R R * A * N S S T N S * S
 14161 - CAACATCCCTATTCTATAGAGACACTATAGAGCTGTGAGTGTAGATTGCGGCATACT - 14220
 - Q H P Y F Y R D T H R A C V V D C G H T
 - N I P Y S I E T L I E P P V L * I A D I L
 - T S L F L * R H S * S L C C R L R T Y L
 14221 - TGTCAGCTATCTTATTACCATGACTTGAAAGAAGTCATTTACATGGCTGTACAGCT - 14280
 - C Q L S Y Y H Q L K E V H L H W L * Q L
 - V S Y L I T I S * K K C I Y I G C N S L
 - S A I L L P S V E R S A F T L A V T A *

FIG. 12 Con't

72/106

- 14281 - GACAAATGTTAAAGACACTATTAGCATAGCAGTTGAGCATCACGGATGATGTTCCAC - 14340
 - D K C * R H Y * H K O L * H H R M M F H
 - T N V K D T I S I S S C S I T G * C S T
 - Q M L K T L L R * A V V A S P D D V P P
 14341 - CTGGTTAACATACTAGTGAGCCGCCACATCACCATCTCACTTAACTTGCGCACACT - 14400
 - L V * H I V S R H T * P S H L I L A H T
 - W F N I * * A A T H D H L T * Y L R T L
 - G L T Y S E P P H M T I S L N T C A H S
 14401 - CGTTAGCTAACCTGTAGAAACGGTGTGATAAGTACAGCAAAGTCATTTGGAGCAA - 14460
 - R * L T C R N G V I S Y S K C Y V C E Q
 - V S * P V E T V * * V T A S V M F A S K
 - L R C D K L Q Q V L C L R A R
 14461 - GAACAGAGAGGCCATTATCTAARGCATGTTAGGCATGCTGTCACTTGGATATA - 14520
 - E Q E R P L S * A C * A W L C H I L D N
 - N K R G H Y P K H V R H G S V T F W I I
 - T R E A I I L S M L G M A L S H F G * S
 14521 - CCCAACCCATAAGGTGTGGAGTTCTACATCACTGTAAACAGTTTTAACATAATATGCC - 14580
 - P N P * G V E F P L H H C K Q F L T Y Y A
 - P T H K V W S Y I T V N S F * H I M P
 - Q P I R C G V S T S L * T V F N I L C Q
 14581 - AGCCACGGCTAAACTCTGCTGTCCTAACCTACACAGTAGCTCTCTAGTGCGGCTATTG - 14640
 - S H R K T C L F Q L P Q * L L * W R L L
 - A T V K L A C S N Y H S S S S S G G Y *
 - P P * N L V L P I T T V A P F L V A R I D
 14641 - ACTTCATAATTCTCTGATGAAACTGCTATTGTCATAGTACTACAGATAGAGCACCG - 14700
 - T S I S D E T V Y L S * Y Y R * R H Q
 - L Q * F L M K L S I C H S T T D R D T S
 - F N N F * * N C L F V I V L Q I E T P A
 14701 - CTACGGTGGAGCTCTATTCCTTCACTTAATGGACTACTTGAATTGACTATAG - 14760
 - L R C E L Y S L H * W H T * D S F E L *
 - Y G A S S I L C T N G I L K I H L S Y S
 - T V R A L F F A L M A Y L R F I * V I V
 14761 - TAGGGATGACATTACGCTTAGTATACGCAAAGTCATCTGATCCTCATRACTCATG - 14820
 - * G * H Y A * Y T R K V H L D P H N S L
 - R D D I T L S I R E K C I L I L I T H *
 - G M T L R L V Y A K S A S * S S * L I E
 14821 - AGTCATAATAAGTCAGCCTTACCCATTATTAACTGGAAACCACGTGATTATCCA - 14880
 - S H N K V * P Y P I Y * M G N Q L I Y P
 - V I I K S S L T P F I K W E T S * F I Q
 - S * * S L A L P H L L N G K P A D L S R
 14881 - GATTGTTAACGCTACTTGGTGGCATTAATACGCCACATCGTAACARTCAAGTATT - 14940
 - D C * R L L G W H * Y S H H R N N Q S I
 - I V N D Y L V G I N T A T I V T I K V F
 - L L T I T W L A L I Q P P S * Q S K Y L
 14941 - TATCAACACTCAACTACGAACTGGAGTGTCTGATCACACATGTTGGCAGATTAT - 15000
 - Y Q Q L Q L R I G V V * Y H T L L A D Y
 - I N N F N Y E * E L S D I T H C W Q I I
 - S T T S T T N R S C L I S H I V V G R L *
 15001 - AACGATAATACTGATAATCACTGATAGCAGCGTTGCCATCTGAGCAAAGAGTGT - 15060
 - N D N S H N H * * Q R C H P E Q R R S V
 - T I I V I I I T D S S V A I L S K E E V F
 - R * * S * S L I A A L P S * A K K K C F
 15061 - TTAGTTCAACGAGACTCTCTCTAAAGAAACCTTGTAGACACAGCAAAGTCATAAAAGT - 15120
 - L V Q Q N F L P * R N L * T Q Q S H K S
 - * F N R T S F L K E T F R H S K V I K V
 - S S T E L P S L K K P L D T A K S * K S

FIG. 12 Con't

73/106

- 15121 - CTTTATTAATAACCGGGTTGACAGTTGAAAGCAACATTGTTGTAGTCAGCTA - 15180
 - L Y * N Y R V * Q F E K Q H C L L V O L
 - F I K I T G F D S L K S N I V C * C S Y
 - L L K L P G L T V * K A T L F V S A A T
 15181 - CTGAAAAGCATGTAGTCGGTTATCTAGCAATAAATGCCAGAGCTGCATGCATAGCTG - 15240
 - L K S M * C V Y L A I N C Q K L H A * L
 - * K A C S A F I * Q * I A R S C M H S W
 - E K H V V R L S S N K L P E A C I A G
 15241 - GATCAGCACCATACACTAAAGATTCCCTGAAGACTGAGACGCCAGCTATGTAAGTTACAT - 15300
 - D Q Q H T L K V P * N * D A S Y V S L H
 - I S S I H * K F L E T E T R A M * V Y I
 - S A A Y T K S S L K L R R E L C K F T S
 15301 - CCTGATTAATGTACGACTCTAACCTACGAAATGGTATCCAGTTGAAACRACARAAGNA - 15360
 - P D Y V R L L T H E N G I Q L K Q O K E
 - L I M Y D S * L T K M V S S * N N K R N
 - * L C T T P N S R K W Y P V E T T K G T
 15361 - CACCATACATATTATTTCTACTAGTGGTCRAAACATTGTTAGGTGGAAACACAGTG - 15420
 - H H L Q I F F L L V V Q N L * V E T Q *
 - T I Y K Y F S Y * W S K T C R C N K H S R
 - P S T N I F L T S G P K L V G G N T V E
 15421 - AAAATAACACATTAARGTTGACAAATGAGGATAACCTTATCATCCAAACAGTTAACAC - 15480
 - K I T H * S L H N E G Y T Y H P N S * Y
 - K * H I K V C T M K D T P I I Q T V N T
 - N N T L K F A Q * R I H L L S S K Q L I Q
 15481 - AATGGGGATGGTAGCTCTGTCGCCCCATATTAAATAACGGTGGAAAGACAAAGTCCT - 15540
 - N W D G M S G P N I * N N G R R D K V S
 - I G M V C L V P I F K I T V E E T K S L
 - L G W Y V W S Q Y L K * R S K R Q S L S
 15541 - CTTCCGTAATAATCATATTTCGCAATCCACATTAAGTGGTTTCCGGAGTCAGCAT - 15600
 - L P * N H I S A N P T * V V L R P C Q H
 - F R K I I F Q Q I P L N K W F C E I S I
 - S V K S Y F S K S H L I S G F A R S A S
 15601 - CCATATGGACTTCAGCAGCCAATGCCCTAGTCAGACTGGGGATGGCCTACGCAATGAGT - 15660
 - P Y G T Q P M P * S K * G W A S A M S
 - H M G L S S Q C P S Q S E D G H Q Q * V
 - I W D S R A N A L V K V R M G I S N E *
 15661 - AATATGAATCCACAATAGGAACCTCCGCAGCTGGTCTACTGTACGAAATCACCGAAT - 15720
 - N M M N P Q * E L R S R L V L L V R N H R N
 - I * I H N R N S A A W C Y C L Y E I T E I
 - Y E S T I G T P Q P G A T C T K S P K S
 15721 - CGTACCACTGGCTCATTAAGATCCCTGATTAATCTAACATCAGTACGCCCTACATGCCCTGCAT - 15780
 - R T S S H * D P D Y L M S V R L Q C L H
 - V P V P I K I L I T * C Q Y A Y N A C I
 - Y Q F P L R S * L S N V S T P T M P A S
 15781 - CACGGCATAGCATGGCAAGATGTCAGACTTAAATGATTGGCTACACGCTCACCTA - 15840
 - H A * H R R I V Q S L I M I G V H A H L
 - T H S I A E L Y S L * * L A Y T L T *
 - R I A S Q N C T V F N N D W R T R S P K
 15841 - AGTTAGCATATACCCGTAAGATGTCAGGATTCCTACGGAAGTCATACCAATCTTCTTAT - 15900
 - S * H I R V R C Q D S L R S H T N P S Y
 - V S I Y A * D V R I L Y E V I P I L L I
 - L A Y T R K M S G F S T K S Y Q S F L L
 15901 - TGAAATAATCATCATCACAGCAATTGATGTGACGAGTATTCTTTATGATCACAT - 15960
 - * N N H H H S N C M * R V F L L M Y H N
 - E I I I I I T A I V C D E Y F F * C I T I
 - K * S S S Q Q L Y V T S I S F N V S Q L

FIG. 12 Con't

74/106

15961 - TACCCCTCATCAAATGAGCTAGAGCATAGACTAAATCAGCCATTGTGTTAGTTAGAC - 16020
 - Y P H O N D V E H R L N Q O P L C I * L D
 - T L I K M T * S I D * I S H C V F S * T
 - P S S K * R R A * T K S A I V Y L V R R
 16021 - GCTGACGTGATATGTGTTACCATGTCACCATCTACTCTAAACTGAAAAGTCATGGA - 16080
 - A D V I Y V V P C H H L D * T * K S H G
 - L T * Y M W Y H V T I Y S K L E K V M D
 - * R D I C G T M S P S T L N L K K S W T
 16081 - CAGCAACCSCTGACATCTTAAACAGTTATAATGCTCTCATGGTAGTTG - 16140
 - Q Q P L D N N L * P S Y K * S L B V G S *
 - S N R W T I F N Q V I N S N L F M L V V R
 - A T A G Q S K L * I V S S C W * L D
 16141 - ACATAGTATGCCCTTAACATAAAAGTAAGGCTAATAATTGCCTCCATCCTCT - 16200
 - T * Y A S * L Q S K S L I N C L P H P S
 - H S M P L Q V K R V * * I A F L I L L
 - I V C L L T T K * E S N K L P S S S F S
 16201 - CCTGGAAGCGCACAGCRAATTAGTTTGTAGGACTTGTGAAACACAGCATTTCCTGTT - 16260
 - P G S D S N * F L G T L Q N Q H F F R C
 - L E A T A I F S * E L C K T S T F F V V
 - W K R Q Q L V F R N F A K P A L F F S L *
 16261 - AAATATCACAAAGCCCTGTAGACGACATCAGTACTGTGCGTGCCGACGGTGAAGAC - 16320
 - K Y Q K P C R R H Q Y * C L C R T V * D
 - N I K S P V D D I S T S A C A A R C K T
 - I S K A L * T T S V L V P V P H G V R R
 16321 - GGGCTGCACTTACACCGCAACCGCTTAAACCGTTGATGCACTGGCAGACTGCATCAA - 16380
 - G L H L H R K P V * K R * C I R R L H Q
 - G C T Y T A N P F K N V D R S A D C I K
 - B A L T P Q T R L K T L M H M P Q T A S R
 16381 - GGGTCCGGAGTTGGTCACAACTCACGCCATACCTTCCACATCCCGACAGGGTACA - 16440
 - G F A E L V T A I T F P H S A D G T
 - G S R S W S Q L Q P F * P F H I P Q T V Q
 - V R G V G H N Y S H N L S T F R R R Y R
 16441 - GACTGTGTTCTAAGTAAACCCACTGGGTCTAGCACACAAGTGTAGGTATTTGGAC - 16500
 - D C V S K C K T H W V I S T S G R Y L D
 - T V F L S V S K P T G S L A Q V V V G I W T
 - L C F * N P L G H * H K W * V F G R
 16501 - GTACTTACCTTCAAGTCACAGAACATCCATTAGATTGGATGTCATGTGCACTACA - 16560
 - V L T F Q V T E S F R I I W M V N V A S T
 - Y L P F K S Q N P L G F G W S M W H L Q
 - T Y L S S H R I L * D L D G Q C G I Y N
 16561 - ATACAGACAACTAGAACGACCAACAAAGACACTGGTCCATGTAGCTCTGGTGTAC - 16620
 - I Q T T * S T T K G L L V H V S F W C Y
 - Y R Q H E A P P K D S W S M L A S G V T
 - T D N M K H A Q R T L G P C * L L V L Q
 16621 - AGTAATTGGCTGTCTGTACAGTGGTGTACACAAACATCTTCACAGTTGGTGTAC - 16680
 - S N C L S C T S V C T Q O H L H T V G D W
 - V I A C P V P V C V H N I F T Q L V I G
 - * L P V L Y Q C V Y T T S S H S W * L V
 16681 - TTGTCTCTTCACTGCTAGSTAATCTTATAGCTTATAGCTTCTGGTGTAC - 16740
 - L S S T C * V I L I C F S R V Y C K S T
 - C P P L A R * S L Y A L A G S T A K A Q
 - V L H L L G N P Y M L * Q G L L Q K H R
 16741 - GAAGGAAAGCACAAGTGGATTGGAGGTACTCTGTAGCATTCAGCTGAGACGCTAC - 16800
 - E G K H S * I G R Y F C S I S S L K T Y
 - K E S T V E L A G T S V A F P A * R R T
 - R K A Q L N W Q V L L * H F Q P E D V L

FIG. 12 Con't

75/106

16801 - TGTAGCAGCTAAACTGCCCCAACCATACCTCTATTAGTTAGCTTAAAGCCCTTGATGAA - 16860
 - C S S * T A O H H T S I * V V * A F D E
 - V A A K L P S T I P L F R L F K P L M K
 - * Q L N C P A P Y P L Y L G C L S L * * S
 16861 - GTACAAGTATTCTACTTCTAGGCCCTTTGGTGTGTCGTGTAACAAACCTACAGGTGGTC - 16920
 - V Q V F H E R P F W C V C C N K P T R W F
 - Y K Y F T L G P F G V S V T N L Q G G S
 - T S I S L * A L L V C L * Q T Y K V V P
 16921 - CAGTTCTGTGAAATGTTACCTGTACCCATCACTTCTAGGGATCTAGGCCATTGAGATC - 16980
 - Q F C V N C T C T I T L R E S S P F E I
 - S S V * I V P V P S L L G N L A H L R S
 - V L C K L Y L Y H H S * C I * P I * D L
 16981 - TTGGTGTCTGATAGTAACTGGCCAGCACACCTCCCTCGAATGTTAGTAGTC - 17040
 - L V V * * C Q H K P T S L R I V I V G
 - W W S D S H A S T N L P P F E L L * * A
 - G G L I V M P A Q T Y L P S N C Y S R Q
 17041 - AAGTGCATTGTCATCAGTACAGCTTGTGIGTGCTGACCGCCACAGGCATCTGTG - 17100
 - K C I V I S T S C L C G T S R T G H L S
 - S A L S S V Q A V C V P R A Q D I C R
 - V H C H Q Y K K L F V W Y Q P H R T S V V
 17101 - TAGTGCCTGACTGACTGATCTATTCTGTAGTTAACAGCTGACTGTGCTCTAGAGC - 17160
 - * C Y W T Q F I I L * F N S * V G S * S
 - S A T G L S S L F C S L T A E L A L R A
 - V L L D S V H Y S V V * Q L S W L L E L
 17161 - TGTAACATAAGGCCAGGCAAATTGGTGAATTGTCCATGTATTTCATAAGTTG - 17220
 - C C N K R P S Q I W * I V H V N F T K L
 - V T I R G Q A K F G E L S M L I S L S *
 - * Q * E A K P N L V N C P C * F H * V E
 17221 - AACAACTTGGCATCGGCATCAACACTTGTGAGATTCCAGAGTGCAGATGCCATGT - 17280
 - N N L A I R I N N N L L D F P E C R C I C
 - T I L L S A S T T C W I S Q S A D A Y V
 - Q S C Y P H Q Q L A L G F P R V Q M H M *
 17281 - AAAGGTGTACCATCACAGTGTCTGTAGGTACCATATCAGGACRACARACCATGAG - 17340
 - K G V T I T S V L V G T I I R D N N H E
 - K V L P S Q V F L * V P * S G T T T M S
 - R C Y H K C S C R Y H N Q G Q Q P * V
 17341 - TTTGGCTGTAGTCATAAGGTATGTGTAGTTGGAAACACAAACCATCACGGCATGT - 17400
 - F G C C S Q W Y D V E W N T T I T T R I V
 - L A A V V N G M M L S G T Q P S R A L L
 - W L L * S M V * C * V E H N H A H C *
 17401 - GATAATGTTAGTCATCTTACAGCTTCTGAGCATAGTGAAGAGCATGGT - 17460
 - D N V K C I I K L P K H S E E H C L
 - I M L L S A S L S F S L S I V K S I V C
 - * C C * V H H Y Q Q A S * A * R A L F A
 17461 - CATAGCACTAGTCACTTGGCCCTCTGTCTCAGATCTGGCTGTGACATTGGGT - 17520
 - H S T S Y F C P L V L R S C L F V H L G
 - I A L V T F A L L L S S D L A C L Y I W V
 - * H * L L L P S C P Q I L P V C T F G S
 17521 - CATAGCCGTACATGCCATTCTTCAACTCTGGCTTCATGGCAGCATCAGGTCAAACTC - 17580
 - H S L I C H L F Q L A L H G S I T V K L
 - I A * S A I F S N L R C M A R A R S R S N S
 - * P D L P S F P T C V A W Q H H G Q T Q
 17581 - AGATTTAGCCACATCAAAAGTTCTTCTTAACTTTGGAGAACGACTCAGAATCACCAT - 17640
 - R F S H I Q R F L * L F P E N D F R I T I
 - D F S L A T F K D F F N P L R T T S E S P L
 - I * P H S K I S L T F * E R L Q N H H *

FIG. 12 Con't

76/106

17641 - AGCTTACAGCTGCTATAGGCCCTGGGCAGTCAGGGCATAGCGGCATATGATGGTAAAGA - 17700
 - S Y S L L I G L L G S G I S G I * W * R
 - A T A C S * A S W A V A * A A Y D G K E
 - L Q P A H R P P G Q W H K R H M M V K N
 17701 - ACTAAATTCTGARGCARTAGCCCTGAAGAGTAGCAGCGTTATCGAGCATTCTCGCACAA - 17760
 - T K F * S N S L K S S T V I E H F L A Q
 - L N S E A I A * R V A R L R L S S I S S H N
 - * I L K Q * P E E * H G Y R A F F R T T
 17761 - CCTTAAATGTCACGCCCCGATGGATAAGCAAAAAGACAAAAGAGAACATCT - 17820
 - P I N V Y S T L H G * Q N R Q K R N H L
 - L L M S T A P C M D S K T D K R E T I F
 - C T G G A A G C T I C A G T G T C T T T G C A A G A G A T A T C A T T G T G G A G T T G A C A C A T T G - 17880
 - L E S F S C V C K K N I I V E L Y T L
 - S K A S V V S F A R R I S L W S C T H C
 - R R K L Q L C L L Q E E Y H C G V V H I V
 17881 - TGCCCACAAATTAGAAGATGACTACTCTAAGTGTGAAAGAACAGACAAAGAGAACATCC - 17940
 - C P Q F R * L Y S K L L K N R E Q Y H
 - A H N L E B D D S T L S C * R T E S S T T
 - P T I * K M T L L * V V E E P R A V P Q
 17941 - AGATGTCACCTTACGTCAGACATTTAGACTGACAGTAGCAACCTTGATACTGGTT - 18000
 - R C A L Y V R H F R L Y S S L D T W F
 - D V H F T S D I L D C T V A T L I H G L
 - M C T L R Q T F * T V O * Q P * Y M V Y
 18001 - ACCTCCATTAACCAAACAATTATGTTAAAGCTTGAAGCATCAATACTCTTAGGAGG - 18060
 - T S N T Q Q L N V K L E S I N T L R R
 - P P I P N Q N L M L S L K A S I L L L G G
 - L Q Y P T T * C * A * K H O Y Y S * E A
 18061 - CAAACCCCTGGAGATCTCATACCTAAATTCTCTGTAGAGACCAAGTAGCTATAAAC - 18120
 - Q K P L G V H I P K F L C R D Q V V I N
 - K S P W E F I Y L N S C V E T K * S * T
 - K A P G S S Y T * I L V * R P S S H K H
 18121 - ACCRAAGAGTAAGCTGAAGTAACGGTTGAGTAACAGAAAAGGCCAAGTAGCAGCAGCA - 18180
 - T K S K P E V T V E * T E K A K V A A A
 - P R V S L K * R L S K Q K R P K * Q Q Q
 - Q E * A * S N G * V N R K G Q S S S S N
 18181 - ACATAATGCTTAAGAACANTAAACRAAGCATGATACACTGTAAGGTGTGCGATATAA - 18240
 - T I A * E T I N K H D T L * G V A S N K
 - Q * P K K Q * T S M I H C K V L P V I N
 - N S L R N N N K Q A * Y T V R C C Q * * I
 18241 - TAACATGGGTAATACTCACACACACACTATAGCTCTAGCTAAACATGATAGT - 18300
 - * Q W V I L N T H K H Y S S S * K H D S
 - N N G * Y S T H T N T I A L A K N M I V
 - T M G N T Q H T Q T L * L * L K T * * S
 18301 - CGTAAACGACACCCAGAATAGTTAGGGTTACAGAAAATAACTAAAGCCACATGGAAATAGC - 18360
 - R N D T R I V R G Y R N N * G P H G N S
 - V T T P E * L E V T E I T K A H M E I A
 - * R H Q N S * R L Q K * L R P T W K * L
 18361 - TTGATCTAAAGCTTACCATAGTAGACTTGTAAACAGTCAATGACATTCTCATCAGTGT - 18420
 - L I * S I T I V D F V N K C N D I H Q C
 - * S K A L P * * T L * T S V M T F I S V
 - D L K H Y H S R L C K Q V * * H S S V S
 18421 - CCAAAACCTCTAGCAGCATCATATAACAGTCGGAGCTGTCATGAGGATAAGCAAAAC - 18480
 - P N T S S S I I I N S A S C H E N K Q N
 - Q T R L A A S S * T V R A V M R I S K T
 - K H V * Q H H H K Q C E L S * E * A K L

FIG. 12 Con't

77/106

18481 - TAAAGCTGAAGCATACATACACAACTCTTAAGCCTATAACAGACAAGCTAGTGTCAGC - 18540
 - * S * S I H N T I L K P I T R Q A S V S
 - K A E A Y I T Q S L S L * F D K L L V S A
 - K L K H T * H N P * A Y N Q T S * C Q P
 18541 - CAACTCGCCATGTCATGATAACGATCACCCAGCTAGCAGGCGATGTAACCATATTAAA - 18600
 - Q F K P C H D T H H P A S R H V D H I K
 - N S S H V M I R I T Q L A G M * T I L K
 - I Q A M S * Y A S P S * Q A C R P Y * S
 185601 - GAAAGCAACTGTTGCAAGAGAGGTAAACGAAACAGCAAGAATGGCTTATGGCT - 18660
 - V S N C C K R R * Q K Q A Q E C V L M L
 - * A T V A R E G N R N R K H K N A C L C L
 - K Q L L Q E K V T E T S T R M R A Y A *
 18661 - AACAAAGCAGCATAGCCTAGCGCAATTGCCATTAACCGAGTAAATGGCAAGAACG - 18720
 - N K Q H S T C S N C H N T K S K W Q E S
 - T S S I A H A A I A I P R V N G K K A
 - Q A A * H M Q Q L P R * Y Q E * M A R K K H
 18721 - ATTCTCGTAACACGAAACAGCTGACCGACTTGACTCTTGAAACAAATCATAGTGA - 18780
 - I L V N K E K Q * P L C T L N K N Q * *
 - F S * T K K N S D H C V C L * T R I N S D
 - S R K Q R K T V T T V Y F E Q E S I V M
 18781 - TGTCAGGAAGTAAACAGCATCCATGATGAGTGGCCCTAAACATTCTTGAACTTAC - 18840
 - C Q E S * K H P M M S A L N N F L E L T
 - V K K V K S I Q * * V P L T I F L N D P
 - S R K L K S N D E C F * Q F S * T Y L
 18841 - TTGGAAAGTAAACACAGAGCATTTGCTAAACACATAATGGTAACTCATCTCTAA - 18900
 - L E G N T R A L S N N I K W C K L I F *
 - W K V T P E H C L T T S N G V N S S S K
 - G R * H Q S I V * Q H Q M V * T H L L K
 18901 - AATAGTGTCTAACAGGATAGTAGGACCATTCATACCATTTGCGAGCAGCTTTCAACG - 18960
 - N S A T K D S T T I T L Q Q L F Q S
 - I V L P R I V R P F I P F C S S S F K A
 - * C Y Q G * Y D H S Y H S A A A L S K Q
 18961 - AGCACACATATCTAACAGCGAAATTCCTGTTGAGCAGAAAGGGTCCCATAATGTCAC - 19020
 - S T H I * D G N S C L S * S R K R S Q Y V N
 - A H I S K T A I P V * A E R G P N M S T
 - H T Y R Q F L R E Q K E V P I C Q H
 19021 - ATGATCTTGTCAAGGTTACATGGTACTTCATGCCACAAAGTTAAAGTCATCAA - 19080
 - M I L C Q R F I V V L H C H K V K V I Q
 - * S C V K G S * L Y F I A T R L K S F K
 - D L V S K V H S C T S L F E Q G * S H S K
 19081 - AGTAGTGGTGRATCTTAAAGAACCCATTCATCCATTGATAACAGCAGCATACGCCA - 19140
 - S S G E S I K K P P I T I D N S S I Q P
 - V V V N L L R N H L S P L I T A A Y S H
 - * W * I Y * E T T Y H K * * Q Q H T A M
 19141 - TGCCAAACACATTTAAATGTTAGGGTTGTCCTGACTCTGCAGCTGTGGCGAGTTGTC - 19200
 - C Q N I * G C V C T C S L C S L S V
 - A K T F N V M V V S V P A A C A V C L S
 - P K H L M L W L C L Y L Q P F V Q F V C Q
 19201 - AACAAATGGACCATAGATTACCTCTAACGTAGTACCCAGCTGTACTCTGTGGAG - 19260
 - N K W T I E F T F * V S T S V Y S C W K
 - T N G P * N L P S K S V P A C T P V G S
 - Q M D H R I Y L L S Q Y Q R V L L L E A
 19261 - CTCCATATGATGATATAGCAGAAAGACCCAGATCATATAATCAAGTTAAACCAACAT - 19320
 - L H M M H I A E R R H A I I I N V K T N T
 - S I * C I * Q K D T Q S * S M L K P T L
 - P Y D A Y S R K T R N H N Q C * N Q H Y

FIG. 12 Con't

78/106

19321 - ACCACATGATCATTAGAAAGAACCCTTATGTATGTTAGGTCATGGCACACTG - 19380
 - T T * S I K E R T F N G M I R S H G T L
 - P H D P L R K E P L M V * L G L M A H *
 - H M I H * G K N L * W Y D * V S W H T D
 19381 - ATAACACCCAGATGGTGAACCATGTTAGCATGCTRAACTGAAAATGTTGACCGGTG - 19440
 - I N T R W * T I V A C * N * K C L T R L
 - * T P D G E P L * H A R T E N V * P G W
 - K H Q M V N H C S M L E L K M F D Q V G
 19441 - GATA CGG CAA ATT TAT CTGGG TGT CTT AGGT TAGA AGT ATCA ACT TT A G C T A A G - 19500
 - D T D K F I L G C L R V R S I N F K P K
 - I R T N L Y L G V L G L E V S T L S L S
 - Y G Q I Y T W V S * G * K Y Q L * R * A
 19501 - CAGACAATTTGATAGAACATGCCAATACACGAGTTGACATGCTGACCGCTGAACAG - 19560
 - Q T I L D H R M A N N T K L N I A S L N K
 - R Q F C I E W P I T R S * T L P A * T R
 - D N F A * N G Q * H E V E H C Q P E Q E
 19561 - AAAGCTGGTTGGAAATGAGCAGATCTTCATAGTTAGGATTAAAGCATGTC - 19620
 - K A M V G F N E Q I F I V R I K H V F
 - K L W L D L R M S R S S * L G L S M S S
 - S Y G W I C E * A D L H S * D * A C L L
 19621 - TGCTGTGCAATGACATGTCCTGGACAGTATAGCTGTATCCARACCAATCCATTAAAG - 19680
 - C C A N D M S W T V Y C V I Q P Q S I K
 - A V Q M T C L G Q Y T V S S N H N P L R
 - L C K * B V L D S I L C H P T T T I H * E
 19681 - AGTTGTAGTCACAGGTACTGTACCATGACCCCTTCACATTGCTGACGGGAATGC - 19740
 - S C S S T G Y L Y H A P F N F A * R E C
 - V V V P Q V T C T M H P S T L P D G N A
 - L * F H R L L V P C T L Q L C L T G M P
 19741 - CATTTCCTAAACACTCTGCRGAACAGCAGAACAGTATTGATGTCITGGGTGTTGTA - 19800
 - H F P K T T L Q N S R S D * C L W W L V
 - I F L K P L C R T A E V I D V C G G G W *
 - F S * N H S A E Q Q K * L M S V V V G R
 19801 - GAGAACATCAGACCTTGAGTTGCTAAAGTCATTAGCCTTGTCAAGTCAGCAAGC - 19860
 - E N I S T * V A K V I * S L C * V A R S
 - R T S A P E L L K S F R A F A K W Q Q A
 - E H Q H L S C * S H L E P L L S G S K L
 19861 - TGCTTCAGCATGCTGTAGTCTAACGCTTCAACTGAAATCTGTACTGTATATA - 19920
 - C F T I A G S I * G S T E I L V L V I *
 - A S R * L V V S K A P L K Y L Y L L Y R
 - L H D S W * Y L R L H * N T C T C Y I E
 19921 - AGCAAGATACTCTTAACTCTGTTAAGTGGCAACRGTCCTCGCTACGGCATTAGGTA - 19980
 - S K I P V I L C K W Q Q C L A T Q F * V
 - A R Y L L Y C V S G N S V S L R N F R Y
 - Q D T C Y T V * V A T V S R Y A I L G T
 19981 - CATTTCCTGTTGACCAAAGGTACACAAAGCAGCTCTCGAAGGTACTAACATGTAAC - 20040
 - H F L V E Q K G T Q S S L E G T K C N
 - I S L L S K K V H K A A S S K V L N V T
 - F P C * A K R Y T K Q P P R R Y * M * L
 20041 - TCCATTAAACATGACTCTTCTCTAGATAGTTGTTAAGAACATGGCAGTGCTCAG - 20100
 - S I K H D S F P K I V V K E P M A V L Q
 - P L N M T L F L R * L L K N Q W Q C F R
 - H * T * L F S * D S C * R T N G S A S E
 20101 - AGAAATACAGAATACATGATGTTATCCAAAAGGCACAAATGGAGAGAACATGGC - 20160
 - R N T E Y I D C C Y P K R H N R R K H G
 - E I Q N T * I A V I Q K G T I G E N M A
 - K Y R I H R L L L S K K A Q * E K T W Q

FIG. 12 Con't

79/106

20161 - AACCCATTGAGGTGACCGAAAGAATGAAACATCATGGTGAATAAGAATGTCRAGTACAA - 20220
 - K P L K V S Q E * N T I I G E I E C Q V Q
 - N H * R * A K N E T S L V K * N V K Y K
 - T I E G E P R M K H H W * N R M S S T S
 20221 - GTAAAAGACTGAGTAGACTCCGGCAGAAAGCTGTAAAGCTGGTACCRAGCAGCTAGT - 20280
 - V K D * V D S R Q K A V S W Y Q T E Y S
 - * K T E * T P G R K L * A G T R Q S I V
 - K R L S R L P F A E S C K L V P D R V * *
 20281 - GAAAGACATCAAAAACAAAGTCATTAAGCAGCACACATGTTGACTCACCAAACAC - 20340
 - E R H Q K Q K C I S S N N M V V L T K N
 - K D I K N K S S A L A A T T W L Y S P K T
 - K T S K T K V H * Q Q Q H G C T H Q K H
 20341 - ACCTGCTGATTTCATANAGTAGTGCGACGACAGTCAACANTATGGCAATRATACCCAC - 20400
 - T S E F H K V V G S T S H Q Y G N N T T
 - R L N F I K * * A A Q V T N M A I I P P
 - V * I S * S S R O H K S P I W O * Y B O
 20401 - AGCCACTACTGAGGCAGCACATCTAACAGCACCCACAGGTGCRCAAGGGAGTAAAGRT - 20460
 - S H Y * S R H I * S T H R L H K R S K D
 - A T T E A D T S K A P T G C T R G V K M
 - P L L K O T S L H K P Q V A Q E E * R C
 20461 - GTTAGCTATGAGATTCTCGCATCACACACAGAAGACTCTGTATGAGCTCTGTAAATG - 20520
 - V S Y E I H R I N T T E N S * * S S V M
 - L A M R F I A S T P Q K T P D R A L * C
 - * D S S H S H Q H H R K L L I E L C N A
 20521 - CTCATTATTAQAGAACCCATCTACACTGGTAGATAGGCCAAATACCTACTCTGACCTTTC - 20580
 - L I I K N P S T T G R * A N T Y F * P F
 - S L L R T H L P L V D R Q I P T S D L S
 - H Y * E P I Y H W * I G K Y L L L T F R
 20581 - CGATGTACCATGTCAGTACAGTACAGCATCRAAGGTGTTACTACTCTAACAGAACCTC - 20640
 - A C T M S T V L S I K S C Y V S N R T L
 - H V P C L Q Y S A S A K V V T T L T E P S
 - M Y H V Y S T Q H Q O K L L L * Q N P P
 20641 - CRGGTAGTGTAGGAAACTGTGTATGATGGARCCATCCATAAGCACATAACGAGTGTCTGG - 20700
 - Q V S V R K L Y D G T I H K H I T S V W
 - R * V L G N C M M E P S I S T * R V S G
 - G K C * E T V * W N H P * A H N E C L D
 20701 - ACAGAAGTCACATAAGAAATAGAACCCCTAGCAAAATTAGTGTCTACACAAATATGGCAC - 20760
 - T K L T I R L T * Q I S V I T I W H
 - R S S L * E I E P S S K L V S * Q Y G T
 - E A H Y K K * N P L A N * C H N N M A Q
 20761 - AGGTTGCCCCATGACATCTTAAATGTCAGTACAGTCAGCAGCAGAACGGAGGT - 20820
 - R F A H S I L K N C T L S S K N A S R G
 - G L P I A S L K I V H S A A R T Q A E V
 - V C P * H P * K L Y T Q Q Q E R K Q R *
 20821 - AGCAAAATCACTATCTCAATGAGTTGGAGGTGTAGCNAATGTGCCCACAGCACT - 20880
 - S K I T L N E F G R C V A N V A N S T
 - A K S L Y S M S L E G V * Q M L P T A L
 - Q N H Y T Q * V W N K V C S K C C Q Q H *
 20881 - AAAAACAGGAGTAGAAATGCAAGAAGTCACCATGTTGCTCTCAGCACAGTACCCGG - 20940
 - K N T R * K M Q E V T I D C S Q H S T R
 - K T R G R K C K K S P L I A L S T V P G
 - K H E V E N A R S H H * L L S A Q Y P V
 20941 - TAAGCCAGGCACTATGAAACCATCTCTCTGTAAATGATAGCAGCTACTACAGGGCAGT - 21000
 - * A R H Y E T N L S C N D S S Y Y R A A
 - K P G T M K P I S L V M I R A T T G Q L
 - S Q A L * N Q S L * * Q L L Q G S F

FIG. 12 Con't

80/106

21001 - TTTGTCATTTGTATGACCAACCGCTGGCTAACCATGCGTCARAAACGAGCATGTTT - 21060
 - F V I F V * T T L A K P C V K T S M F
 - L S F L Y E P P R W L N H A S K P A C L
 - C H F C M N H H A G * T M R O N Q H V Y
 21051 - ATTTGCAAAACAATCATCAGTAGAATGATGTCACCGAGTGACACCACCTCGATGGCTT - 21120
 - I C K T I I S R N D V T S D T I L N G F
 - F R A K Q S S V E M M S R V T P S * M A L
 - L Q N N N H Q * K * C H E * H H P E W L C
 21121 - GTAAACCAATTGATTTCATTCTGTGTAACCATCATGATGTTGAAATGTATGACTGGCATNAC - 21180
 - V T N N D F I C V T I M D * Q C M Y W H N
 - * P M I S F V * P S W I D N V C T G I T
 - N Q * F H L C N H H L T M Y V L A * R
 21181 - GATAATAACAAACCAATGCAGCAAGAACGCACATAATAATGTCGGCTTAAGGCATAAGTTTAA - 21240
 - D I T N Q C S K N A Q * C G L K H K F K
 - I * Q T N A A R T H N V A L S I S L K
 - Y N K P M Q Q E R T I M W P * A * V * N
 21241 - ACAGACTAACAAATCTTACCAACCCCTGAGTAGGTTTGTAGTTGACATGACACAC - 21300
 - T S T N N L T T L E * D F S S Y D I D N
 - Q V L T I L P P L S E S I L V V M T L T T
 - K Y * Q S Y H P * V R F * * L * H * Q P
 21301 - CGTCTCTAGTGTAGCACAGTGTAGTGTAAAGGTATGTTCTCTTGGCAGCAGTACG - 21360
 - L S S C S T S * C K R Y V V V L L G S S T
 - C L V V A Q V S V K G M L F F L R A A V R
 - V * H K L V * K V C C C S S W Q Q Y E
 21361 - AATTGTTTACGCCAGCTGTTCAGATAAAGCATGATGCTCTTACATTCCAGATGAGTGA - 21420
 - N L F T Q L F R * R H V V F Y I P D E *
 - I C L R S C S D K D M * S F T F Q M S E
 - F V Y A A V Q I K T C S L L H S R * V K
 21421 - AACATTGTGACTCTTGTCTACTTGGGCATITGATAIGCCTGCAATTACAGTCATAACATGC - 21480
 - N I V T F C Y L G I D M P C I T V N T C
 - T L * L F A T W A L I C L A L Q S I H A
 - H C D F L L L G H * Y A L H Y S Q Y M R
 21481 - GCCAAAGATCTCAGGGCGTATGTTTCAACCTTATATAGGTGAGCATGAAATTGTACR - 21540
 - A K I S G R V Y F N L I I G E H E I V T
 - P R S L G R V M F S T L L * V S M K L L Q
 - Q D L W A S C F Q P Y R * A * N C Y N
 21541 - ACTGTCACCTGTCACTTCAACTGAGGTGATGTAAGGTGAGCATTCATAACATC - 21600
 - T V T C H F * V R V M * E F E T F N N I
 - L S P V T S K S E * C E S L R H S I T S
 - C H L S L L S Q S D V K V * D I Q * H P
 21601 - CTTTGTGTCACATGGTATCAGACACACCTTGCGGGCAGCTGACRCGAATGTAGAAG - 21660
 - L C V N I G I N N T L S G S * H E C R K
 - F V S T S V S T T P C R A A D T N V E R
 - L C Q H R Y Q Q H L V G Q L T R M * K G
 21661 - GACACCACTTAARGCTAACCCCTTGTCTACTCGCTGTGAGCTGTACACAGTGCCCT - 21720
 - D T I * S Y T L C * L A V S C S N K C L
 - T P S K A T P F A N S L * A V A T S A L
 - H H L K L H P L L T R C E L * Q Q V P *
 21721 - ARGTTTTCCATAGGAACACTAAAGGTGCTGAAAGGTGTCGACATARGCATCAAACAT - 21780
 - K F F H R N T K S C * K G V D I S I K H
 - S P S I G T L K V A E K V S T * A S N I
 - V F P * E H * K L L K R C R H K H Q T S
 21781 - CTAAACGGAAACTTCAGTACTATCTCCRAAGCTTGATACAAGCCTGGTCAGACAG - 21840
 - L N G N F S T I S N V * Y K S L V K Q Q
 - L T E T S V S L S P T E F D T R A W S S N R
 - * R K L Q Y Y L Q R L I Q E L G Q A T E

FIG. 12 Con't

21841 - AATAGTTGGCACATCAGCTGACTGTAGTACACAGAACGAGACTTAGAAGCAGACTCGTC - 21900
 - N R L A H Q L T V V H R S R L R S R L V
 - I G W H I S * L * Y T E A D L E A D S S
 - * V G T S A D C S T Q K Q T * K Q T R R
 21901 - GCAATTGGACTTGCCATAAAAACTATGACATTAACTAGGCAGTGAACCTTACTGTGCTT - 21960
 - A F G L A I K N Y D I N R Q * T F S V V
 - H L D L P S K T M T L I G S E P L V L L
 - I W T C H Q K L * H * A V N L * C C *
 21961 - AGCTCTCAAATTGCTAAATTGACRAARTGGAGAGCGGATGTCCTCATAGTCCTTTC - 22020
 - S S Q I V * I D K M G E R M S L I G L L
 - A L K L S K L T K W E S G C L S * V F *
 - L S N C L N * Q N G R A D V S H R S F D
 22021 - ACCAGCCTGTCRAAGTAGAGGTGAGGGCGCCATTTCACAGCARCATAACAAAT - 22080
 - T S L V K V E V K R A I F H S N T I N N
 - P A L S K * R * S A P P F T A T L L S T I
 - C P C Q S R G E A R H F S S Q O H Y Q Q Y
 22081 - ATACCGATGACTGCTAGTGGGTGATGGCTTAACTGGACTGACAAATCAGAGC - 22140
 - I R * L V S R V D W S F K L E * Q I T S
 - Y D D W S V G L I G L L N W S D K S R A
 - T M T G Q * G * L V F * T G V T N H E Q
 22141 - AACCTCATCTAAATGATGATCTACCACTGCAAAATGTCACAATTGAGAACATTCA - 22200
 - N F I T N E C T T S A K C V T I E T I P
 - T S S L M N V L P V Q N V S Q L R Q F Q
 - L H * * M Y Q C K M C H N * D N S N
 22201 - ATTGTGAGCTTTCGAGGCCAGGGCTCCATTGCTCATAGACATAGAAAGATCTTCAT - 22260
 - I V S L A E A T A S I C I D I E R S L H
 - L * V L Q K P R P P F A * T * K D L F M
 - C E S C R S H G L H L R R H R K I S S C
 22261 - GCCATTAACATACTAGTGTACACTCAACGGCTGTGGCACGGTGCCTTATAGCAGCAT - 22320
 - A I N N S C T L N A C G T I A L I A H
 - P D T I V V H S T R V A R L R L * H I M
 - H * Q * L Y T Q R V W H D C A Y S T S C
 22321 - GCAAGTCAGAGGGTCAACATCCATGATAGACATAGCTCTCCATATGTAGTAA - 22380
 - A S R R G A T I H D M N I A T L F H M V V E
 - Q V E E V Q P S M I * T * K
 - K S K R C N H P * Y E H S S S I C S R K
 22381 - AGAAGCAAGAACGATGTACATCCTAACCATTCGAGAACGGGTGCATTGTACAAACT - 22440
 - R S K E D V H F N H C R N G C H L Y N T
 - E A K K M Y I L T I A E T G A I C T I L
 - K Q R R C T S * P L Q K R V P F V Q Y *
 22441 - AATGATAACCACTGAGCCAAGAATTGCTGATGAAATGACTAGCRAAATAGCCAAAGAA - 22500
 - N D K P H E P R I A D E M T S K I A K E
 - M I N H M S Q E L L M K * L A K * P K N
 - * + T T * R K N C * * N D * Q N S Q R T
 22501 - CACCTGCAATTAGCTGAAAGACCTAAATAAATAAAAGAATTTCGTGACACATATATGC - 22560
 - H L H Y S * K T * I K E F C E Q H I C
 - T C I I A E R P N K * K N F V V N N I Y A
 - P A L * L K D L I N K R I L * T T Y M P
 22561 - CAAACCCACTAGCGGCCAGACCTAAATGTCAGTCTAGCTTGACGATGAAATCGT - 22620
 - Q N P L S G Q T * N C Q V * L V R * N R
 - K T H S R A R P K I V K S S L Y D E I V
 - K P T Q R P D L K L S S L A C T M K S S
 22621 - CACCTGAAATGTTCAAGAGCTGGATAGAACTGAGGGAGCTTATCCTAAACAAAT - 22680
 - H L N G F K S W I R I K G V * S T * T N
 - T * M V S R S A G * E S R E S N P L K Q M
 - P E W F Q E L D K N Q G S L I H L N K C

FIG. 12 Con't

82/106

22681 - GCTGCAAGGAAAGAACCTTCACAGRAATCCATAGTAGTAACGTTAGACGAATTAGATA - 22740
 - A A R K R T F T E I H S S N V R R I K I
 - L Q G K E P S Q K S I V V T L D E L R Y
 - C K E K N L H R N P * * * R * T N * D T
 22741 - C A T T C T A C G C C A T T A C A T T A A G A G G A G C A C C A A A T T G A T A A G G A C T A C R C C A A A - 22800
 - Q F S N A I T I R R S T K I R * E Y T K
 - N S L T P L Q * E G A P K L D K S T P K
 - I L * R H Y N K K E H Q N * I R V H Q K
 22801 - A G C A C A G G T T A C A C G A T T A G A G A C C T T A A G C A A A T A C T T A A C A C A T A G C C A C A T A G C C - 22860
 - S S S Y T D * R T * A N T * Q Q * P H S
 - A A V T Q I R E F K Q I L N N N S H I A
 - O Q L H R L E N L S K Y L T T I A T * R
 22861 - G A T T C T G A C A C T T A C A T T A G A R A N T T G G G T G A C T T C A C A T A T T A T G C C C A T C C A A A C A - 22920
 - D C E Q F R K F G * L H I I N A G I Q T
 - I V N N L E N L G D F T * L M P A S K H
 - L * T I * K I W V T S H N * C R H P N I
 22921 - T A A T T A G C A C A C T T A C A T T T T T A G C A T A G T T G T A G G T A G T G A A G C T C T A A T - 22980
 - * F S N T L N T I F S N S C R * * S S N
 - N L A T L L T L F L R I V V G S E A L I
 - I * Q H S * H Y F * Q * L * V V K L * F
 22981 - T C T G A A T T G G T A C T T T A C T T A A A G T A C A C A T T G G A C A T A T A T G T A A C A C A T A G G - 23040
 - S R I G T F S K S T Q L E Q * C K H I R
 - L E L V L L V K H N W N N N N V N T * G
 - N W Y F * * K Y T I G T I M * T H K A
 23041 - C A T A T A T T T T A A C A C A C G G T G C T A A T C T C T A G C C P A T T G A T G T G T A A T G C - 23100
 - H I I V K H T L C * S L S A I * C C N C
 - I * L L N T R C A N N L L A Q F D V V V I A
 - Y N C * T H R V L I S * R N L M L * L L
 23101 - T G C T G I C T A A G A T G G T T G A C A T A A G G C C A A A A T T T A C T C C A A G G A C A C T A T T A A T - 23160
 - C C L S * E W F D P I S Q N F T P R N T I N
 - A C P K N G L T * A R I I L L Q G T L L I
 - L V R L M V * H E P K F Y S K E H Y * L
 23161 - T G C A C C A T A C C A T G A T G T G C A A T T G T T T A A C C T A A G G C T A G T G A A A G C T C A T T A G G - 23220
 - C S N T M S G N C F * T * G * * K L I R
 - R A I P * V A V I F V K F P K A S E S S S L G
 - Q Q Y H E W Q L F L N I R L L V K A H * V
 23221 - T T T C T A A T G G T A T G T C T G T T T C C A T A A G C A G G C T A A G A C T C T C A T G C C T A A - 23280
 - F F L N G N A C V C F H I S S H K I L M T *
 - F L M V N M L V F S T * A A I R S S * P N
 - S * W * C L C F P R H K Q P * D P H D D L T
 23281 - C T C T G I G I G T A C T T I A C A C C I T C A T C T G T G T T T A A G T A T G A C T T G C C T A C A C I T C - 23340
 - L L C Y F N A T T P F S S D G L L S M T L P T T S
 - S C V T L T P S S D G L L S M T L P T T S
 - L V L L * H L H L M V * V * H C L Q L R
 23341 - G G T A G T T T C A G C G T C A C A C T C T A G C T T C C T G T A T G T A G G A T T T C C A C T A C T C - 23400
 - G S F H V T L Y D F L L Y G R I F Y F
 - V V F T S H S M T S F C M V G F S T T S
 - * F S R H T L * L P S V W * D F P L L L
 23401 - T T C A G A G G T G G G T T G C A T T I C A C A G C A A G A T T G T C C A T T C C T G T G T C T C A C - 23460
 - F R G G L L T F T S K I V H S L C V F Y
 - S E V G C * L S Q A R L S I P C V V S S T
 - Q R W V V D F H K Q D C P F L V C L L D
 23461 - T G C A G A A C T C A A R T T G A A G T A T C T A C T G G C T T G T A C T C C A A A G A C A C G T A A - 23520
 - C Q N F K * I * S I Y W L C T P K T T *
 - A R T S N E F E V S T G F V L Q R Q R K
 - P E L Q M N L K Y L L A L Y S K D N V N

FIG. 12 Con't

83/106

23521 - ACACCCAAGTGTCTGGTTGAACTGGTCTCTGGCTGACTGGCTGGTTAATGTGCCAACAAAT - 23580
 - T P S V W F E R C L G C S L V N V P N N
 - H Q V F G L N V V L V V A W L M C Q T I
 - T K C L V * T L S W L * F G * C A K Q L
 23581 - TGGCCTATGCAGTAATTACCGCCCTTCCTGAAACTCGCTGAATAGTGCTATACTCGAT - 23640
 - W L M Q * F S T F L E T T R * I V S I V N
 - G L C S N L A P F L K L A E * C L * S I
 - A Y A V I * H L S * N S L N S V Y S Q *
 23641 - AGCCATACATGCCATTCACTGGAAAGATGTGACAGATAGCTCTGCTGAGCTGG - 23700
 - S H Y I A I Q V W E E C D R * L S * S W
 - A T T S P F K S G K N V T D S S R E A G
 - P L R H R S S L G R M * Q I A L V K L A
 23701 - CTTTGAGACCTGTCATTGTTAAATCATCGCAATTITGGTGTAGAACATGTGAG - 23760
 - L C E A C H L I * I I S K F C V R T C E
 - F V C P V I * F K S S A N E V L E H V S
 - L * S L S F D L N H Q Q I L C * N M * V
 23761 - TTGAAATTATCRAAATCTGCCATTGGTAATGGTAGTTGGTACAAGGCTATAGGGCT - 23820
 - F E I I K T R I W * W L S W Y K V Y R L
 - L K L S K L A F G G N G * V G T R S I G C
 - * N Y Q N S H L V M V E L V Q G L * A A
 23821 - CTCCTGATATAAGCATTCCTTTATATAACCCATCCMATTGGTCAATCTGT - 23880
 - L C I V S I I L F I I P I Q F W F N L C
 - S V * * A L S F L * Y P S N F G S I S V
 - L Y S K H Y P F Y N T P I L V Q S S L C
 23881 - GTAACTCATGACTTTATACGACACAGGCTTGTAGTTGTAGTGTAAGATGTTTC - 23940
 - V S N S I E F I R H R L D G C S V R C F
 - * V T P S S L Y D T G L M V V V * D V S
 - K * L H R V Y T T Q A * W L * C K M F P
 23941 - CTGTAGAACATCGTCAGTCAGTCAGTGTACATCTGACATCTTGTAAGGTGAGCTCC - 24000
 - L V E N I S H W P F V L * H L C K V S S
 - L * K T S V T G P L Y S D I F V R * A P
 - C R K H Q S L V L C T L T S L * G E L R
 24001 - GTCATACGATAGGGCTCTCTAGCAGTTATAGTGTAATGACCAACTGTAGTT - 24060
 - V N T I E G L L S S Y M S V M T T L I V
 - S I R * R V S L A V I * V * * P H * * L
 - Q Y D R G S P * Q L Y E C N D B T D S Y
 24061 - ACCAGTGTACTCATCGACATAGAATGTCACCTGCTGTAATTATACGAGGTGG - 24120
 - T S V S L I R T * E C T L L * F I L S R W
 - P V Y S F A H K N V P C C N L Y S A G G
 - Q C T H S H I R M Y L A V I Y T Q Q V V
 24121 - TGCAAGACATCATAACAAAAGRAGACTCTGTGTACTAGATATTGTGTAGCATCAAGGCC - 24180
 - C R H H N K R R L L L Y * I L C S I T T
 - A D I I T K E D S C C T R Y C V A S R P
 - Q T S * Q K K T L L V V L D I V * H H D H
 24181 - ACACACACATCGAAATGNAACACCTGCTTAAAGATTATCATAGATAGACTACCCATATA - 24240
 - T H T W N G N T C L K I I R * S T H I
 - H T H G M E T P V L R L S * D R V P I Y
 - T H N E W K H L S * D Y H K I E Y P Y T
 24241 - CATCACAGCTTACACCCGTTAACGGTAGTAGTTCTGACCACAATCTTACACRCCAC - 24300
 - H H S F Y T R * G S S F L T T M F T H H
 - I T A S T P V K V V V F * P Q C L H T T
 - S Q L L R P L R * * F S D H N V Y T P H
 24301 - ATTAAGAACACTGCGTTTGCAGATTCCAAATTACCATCTGAGAGATGGCTATAGTTTC - 24360
 - I K N S L C R F Q I S M L * K M G H S F
 - L R T R F A D S K L A C C R R W V I V S
 - * E L A L Q I F N * H A V E D G S * F L

FIG. 12 Con't

84/106

- 24361 - TCTGACATCACCAGCTGCCAACAGTTTATTACTGTAAGCGAGTATGACTCCACAAA - 24420
 - S D I T K L A N S F I T V T S E Y E C T K
 - L T S P S S P T V L L L * A S M S A Q K
 - * H H Q A R Q O Q F Y Y C K R V * V H K S
 24421 - GTTAGCAGCATCACAGCACGGCTCTATAATAAGGCCTTGAAGTGCTGGTCATGAA - 24480
 - V S S I T S T G S I I S L L K C W C I E
 - L R A S P A R A L * * A S * S A G A L N
 - * Q H H Q E G G L Y N K P L E V L V H * I
 24481 - TTGACTTCAGCTGTGAAGCTTAATAAACACTAGACAATAACATTGTTTCAGC - 24540
 - F D F K L L K C * * N T R Q I T V I S
 - L T S S C * S A N K T L D K * Q L L S A
 - * L Q R A E V L I K H * T N N N C Y Q P
 24541 - CCATTAAATTGAGTTAACACCAACTGAGGAAATTCCATTTCGTGGTTTAA - 24600
 - P F N * S * T T N L R K F P F L C C V V *
 - H L I E V K P P T * G N F H F F V W N F K
 - I * L K L N H Q L E E I S I S L C G L K
 24601 - AGCAGACATGACTTACCAAAGAAAATCTCATCAAGAGTATGGTAGTACTGRAAGCTC - 24660
 - S R H V P T K T L I K S M V V L E S F
 - A D M Y P L F R K L S S R V W * Y S K A S
 - Q T C T Y Q B N S H Q B Y G S T R K L H
 24651 - ACTAGCTAGTGTCTACATCTAGTAGTACAAGAAGCTTACCCATGATTTACATC - 24720
 - T T * C V I T R * Y K E S L T L M I Y M
 - L R S V S S L G S T K K V L P S * F T *
 - Y V V C C H * V V Q R K S Y P H D L H E
 24721 - AGTTTAAATTGTTGACATCACCATCCRAGTAGTTGGACCAAATCTGCTGTCCATA - 24780
 - R F N F C N I S T I Q V C M W T K L L S . I
 - G L F I F V T S A P S K Y V G P N C C P Y
 - V * F L * H Q H R P S M L D Q T A V H M
 24781 - TGTCATAGACATACCAACAGCTGTGIGTGGAGATAGTGTGTCACAGTGTGAAACAC - 24840
 - C H R E I H K C V E I V H S C E H
 - V I D I S T S C V W R L L V L S T V V N T
 - S * T Y P Q A V V C G D * C C P Q L * T L
 24841 - TTTATAGCTTAACCTCCGCAAGGATAAGAGACTCTTAGTTGTCRAGTGAAGAAC - 24900
 - F Y S L N L P Q G * E T L * F V K * K N
 - F I V L T S R D K R L D F S L S S E R T
 - L * S * P P A G I R Q D S L V C Q V K E P
 24901 - CTACCGTCAAGATGACAATCGAGGGCTCTCACAGAGTAGTACACAAATTGTCACC - 24960
 - L T V K M K L D G A L Q S V V H N F V T
 - S P S R * N S T G L S R V W Y T I L S P
 - H R Q D E T R G S P E C G T Q F C H H
 24961 - RCGCTTAAGAAATTCAACACCTRAGCTCTGCTGCTGTGAATAGGACCAATCTGTGTA - 25020
 - T L K K F N T * L C T L S * I G P I S V
 - R L R N S T P S V R C P E * D Q S L *
 - A * E I Q H L T L Y A V L N R T N L C K
 25021 - AGAGCCGCCAARGAAACTGTTCTACAAAGTGTCTCACAGATGCTTGTAGCAGGAAT - 25080
 - R A S Q R N C F Y K V L L R C L * * R S
 - E F P A K E T V S T K T C S S D V F D D E V
 - S Q P K K L F L Q S A P Q M S L M T K *
 25081 - GAGGTATCCATTATGAGTAAACGACATCGGTGTAGTACTGACACTACGGCAGGAGC - 25140
 - E V S I I C S M S I W * Y * H Y G R S
 - R Y P L Y V V T A S G D D T D T T A G A
 - G I H Y M * * Q H L V M I L T L R Q E L
 25141 - TTAAAGAGAACCCATACAGGGCCAGGCCCTTCAGAGATTAACCAATGTCACATACCC - 25200
 - F K R T H T A R S L F K I K T M C H I T
 - L R E R I Q R A A S S R L K P C V T * P
 - * E N A Y S A Q P L Q D * N H V S H N Q

FIG. 12 Con't

85/106

25201 - AATTGGCATTGACAAAGCGGCTATTTAGAGAGTTCAGCTTCGTATAATAGAAGCTAC - 25260
 - N W H C D K R L I * R V Q O L R N N R S Y
 - I G I V T S G S F R E F S F V I I E A T
 - L A L * Q A A H L E S S A S * * * K L Q
 25261 - AGGCCTTACTAGTATAAAGAAGATCGGAACCCATAGTCACAGATGCCCTCTGAAT - 25320
 - R L F T S I K E E S D T I V N D A L L N
 - G S L L V * K K N R T P * S T M P S * I
 - A L Y * Y K R R I G H H S Q R C P L E F
 25321 - TTAAATTCTTATACCTACGGTGGATGCCATTATGGCTCTAACATCCATGCTAT - 25380
 - F N S F I L T L D G C H Y G S N I H A Y
 - L I P L Y L R W M V A I M A L T S M H I
 - * F L Y T Y V G W L P L W L * H P C I *
 25381 - AGGCATTAATTTCTGTCCTTCAGCATGAGCAAGGATTCCTCAATTCTCAGGATAC - 25440
 - R H F S C L F S M S K H F S Q I P G Y
 - G I N F L V S S A * A S I S L K F Q D T
 - A L I F L S L O R E Q A F L S N S R I O
 25441 - AGCTCCTGAGATCTCTCCCTAGCATTAAGGTGCTCTGAAAGGTAGTACATARAATGCAGA - 25500
 - S S * N L F L S I R C F * R * Y I K C R
 - V P R I S S L A L G A S E G S T * N A D
 - F L E S L P * H * V L L K V V H K M Q I
 25501 - TTGCAATTCTTAAGAGCACTTAGCTTCAGGTGTATACCCAGCACATCCTGTCC - 25560
 - F A F L K C S S L S F L K C I T S T S L S
 - L H F L R A V L A S S S V * P A H P C P
 - C I S * E Q S * L P Q V Y N Q H I L V Q
 25561 - AGGTAAGTGTATACATCATCAACTGCACTTCTCAAGGCTCTGAGAGCATCTC - 25620
 - R V R G Y I L I N W H F L L Q S S * E H L
 - G Y V V I Y S T S G T F F K A L E S I S
 - G T W L Y T H Q D A L L S S K L L R A S Q
 25621 - AGTAGTCACCCAGCCTTTGGAGGTATACRACAAGTGTATCACCACTAGTGT - 25680
 - S S A T S L F G G Y Y N T S D I T T S D
 - V V F P P A F L E G I T T Q V I S P L V I
 - * C H Q P F W R V L Q H K * Y H H * * *
 25681 - AACATCACCAACCATGTAAGGCGATCCTCTCAAGGAARGACATCTCCACTCTARG - 25740
 - N I T Y H V R C I L L K E R H I F T S K
 - T S P T M * G A S F S R K D I S S P L S
 - H H L P C K V H P S Q G K T Y L H L * A
 25741 - CATGTCGAGATCATGTAAGACTACCATGATATCAGAACACAGAGTAACCTATT - 25800
 - H V L R I M V K K L T I D I S K Q E * L I
 - M F * E S W * S L P L I S A N K S N L L
 - C S E N H G K R A Y H * Y Q Q T R V T Y W
 25801 - GGTARGARCTTAGTTCTCAGTGTGGTAACCTCATCAATGCAGGCCCTTATTTT - 25860
 - G K K L S F F Q C C G N D I N A G L N F
 - V R N L V S S S V V V T S S S M Q A L I F
 - * E T * F L P V L W * P H Q C R P * F L
 25861 - TGGCTTCACATCGACAGGCTCTGAGCACAGATTCTCCTCAGTTGGATCTCTGT - 25920
 - W L H I D R L L Y D R F L L S F G I F C
 - G F T T S T G F C T T D F S S V L E S S V
 - A S H R Q A S V R Q I S P Q F W N L L C
 25921 - GTTGGTGGCTCTCTGTGTTAGGTGCTCAGCTAGGCTCAGGTTATCAAGATAATC - 25980
 - V W W L L L F R C F H S R L Q V I K I I
 - F G G S S C L G A S T L G F R L S R * S
 - L V A P L V * V L P L * A S G Y Q D N P
 25981 - CATGACAACTCTCTCATAAAGGCTTGTGCTAGTCAATATAACCTGTAGAAC - 26040
 - H D N L L L I K S F V I D C N I N L C T N
 - M T T C S * R A L S L T A I * T C V R T
 - * Q P A H K E L C H * L Q Y K P V Y E P

FIG. 12 Con't

86/106

26041 - CGTCCTGACGCCAACACTGTAAAGCTGAAAGTTGGTTAGCACCATAATGCCTGCTGACCAA - 26100
 - R L H A H L * R L K W F S T K Y C A C * Q
 - V C T H T C K C D * S G L P A M N P A D N
 - S A R T L V K T E T V E * H Q I L T L T
 26101 - CAAATGGTCTCGAAGTAAAGCTCTGCTGATGATTGAGCTTGAAGAGTC - 26160
 - Q W C K * D V L * I E I F I C C L K K L
 - N G A S K M S C E L K F S Y A R L R S W
 - M V Q V R C P V N * N F H N L P * E R E G
 26151 - GATGTCCTCACCTGCTTAGGTTAGGTCCAAACATGCAGACACTCTTAGCAGATT - 26220
 - D V L T C I * V R S S N N M O T L L S R I
 - M S S P A F R L G P T T C R H F L L A R L
 - C P H H L L G * V Q Q H A D T S * Q D Y
 26221 - ATGTCCTAGAACAGCAACAGACCCCTCTACTGTAAAGAGGCCATTAGCTTATGTC - 26280
 - M S R K P S V R L * E C H L R * L N V I
 - C P E S K Q D P P T V R G P F S L M * S
 - V Q K A N K T L L * E G H L A * C N H
 26281 - ATCACTCTCTTGTGATGGCACCATTTGGCTGCTTGTGATGCGACCTCTACACCC - 26340
 - I T L L E L H G T I G C L V E C T C Y T T
 - S L S F C F M A P L V A L L S A P A T P P
 - P S P F C V H W H H W L L * V H L L E R H
 26341 - ACCATGTTTCAGGTGTTATGGTACGCACTTACAATCCATAGGATTAGCCTTGTGGC - 26400
 - T M F Q V Y V S S I Y N N H H R I S T L C
 - P C F R C M L A A F T I T I G L A L C A
 - H V S G V G C * Q H L Q S P * D * H F V P
 26401 - CTCCCTTAACGATGTCAACATTTATGCGACATTTGTCAGTAGTTTAATAACCGT - 26460
 - L L D N V T T F N G N I * T T T I T S
 - S L T M S T H L M A T L S V S F K * P V
 - P * R C Q H I * W Q H C Q C * V L N N O *
 26461 - AACATGTTAACCTGGTTCTCAGGTGTTAGGTTCTGGTTCTGGCTCAATCTGTGTC - 26520
 - K L L I N W F P R C R F W F W L N L * E L
 - N * L T G S S G V G G S G S I S D C S
 - T D * L L L Q V * L V L M A T L S V S F K * P V
 26521 - AGTAGTATCATCAGGCGATCTCTCTCTTCTCTCTCACTCGAACATTGCTGAGTGTGCA - 26580
 - S S I Q P V F L F F F L N S N C F S *
 - V V S S S Q S S S S S S S S T R T V S A B
 - + Y H P A S S L P L F P Q L E B L F Q L R
 26581 - GGCACCAAATTCAGGGAGACCTTGATATCATCTCTGACCGFACTCTGTTCA - 26640
 - G T K F Q R E T L I I I L C T V L M F T
 - A P N S R G R P * * S S S V P Y S C S Q
 - H Q I P E G D L D N H P L Y R T H V H R
 26641 - GGTGTTCATCAATTCTCTCTCCACACTGCGATCTCTCTCTCTCTCATCTGGGG - 26700
 - G F I N F F F L T L C I V L F F L I W R
 - V S S I S S S S S H S A S S S S S S S G G
 - F H Q F L L P H T L H R P L L P H L E G
 26701 - GTAAGGAGAACATAATCATCGTGTGAAAGTTCTCTCACCGACATCAATAAGTA - 26760
 - V K G T I H T * * K V F F T S I I K * V
 - * E O Y I R D E K F F S S P A S S N K *
 26761 - GAAATGTTACACTCTCACCTCAAGTCATACCCATGTTGGTAAAGGAGTCAGAAC - 26820
 - E C S Y T P L I K I N T H V G K E I R N
 - N V A T L H S S R S I P M L V R R S E T
 - M * L H S T H Q D Q Y P C W * G D Q K L
 26821 - TGGTGTGTTAGTCTTCACACAGCCTCTGACACATGCAACATGCAACATGCTT - 26880
 - W L * C S C L C Y N T S N F G
 - G C K V F T T A S A T T H A N S V T S V
 - V V K S S C O P F L L O H N O T O * I R Y

FIG. 12. Con't

87/106

26981 - ACCGGATTCAACAGTGTAGACAGAGCACTTTCATTRAGCAGCTTGTCACACGTTCACTC - 26940
 - T G F N S V D R A L F I K H F V N T E I
 - F D S T V * T E H F S L S T L S T R S S
 - R I Q Q C R Q S T F H * A L C Q H V H Q
 26941 - ARGCTCAATGATGATCTCACATCTCTGTAACTCTGACTCTCCAAACAGTATCTCTCC - 27000
 - K L K C D S I L V T L N F P N S I F S
 - S S N V I L T F L * P * T S Q T V S S P
 - A Q M * F S H S C N L E L F K Q Y L L Q
 27001 - AAAAGTTAACCTTTAATGGTGCACCCCTTAAAGCGAAAGACATTTGTTGAGGCCAG - 27050
 - K G Y T F N W C T P F * A K D I V C S Q
 - K V T P L I G A P P F K R K T L F V A S
 - R L H L * L V H P L L S E R H C L * P V
 27051 - TAACCCAGAGACATGCGCATTTGCTCTTCTCTTAATCTCTAAGAGCATGAGCC - 27120
 - * T R R Q C A V L F F V L N L * E H E A
 - K P G D N D N A Q Y C S C L S L I S K S M R P
 - N O E T M R S I V L C P * S L R A * G H
 27121 - ATTACACAGACTGGTGTGGCAAGGATAGCTCCATTGTAAGCTATCAA CGGGCGTCTC - 27180
 - I Y T D W C A D D S S I C E A I N G R L
 - F T Q T G V G P T I A P F V K L S T G V S
 - L H D V R * L H L * S Y Q R A S R
 27181 - GAGTCCTTCAGTTACACGGTCTTGAGAACACCTCTCAGGGTAAGTAGCTGTGTCATG - 27240
 - E C F E F T V L E N N L R G K Y C V M
 - S A S S S P F L R T T S S E V S T V S C
 - V L R V H R S * E Q P P Q R * V L C H V
 27241 - TGAATCACCTTCAGAACAGTACTCTTTGGCTTAAGAGGCATGAGTAGTTGCRG - 27300
 - * I T F K K G Y F F W C L K R H E * L Q
 - E S P S R R K V T S F G A L R G M S S C S
 - N H L Q E R L L L L V P * E A * V V A A
 27301 - CTGCTCTTCAGCTACAGTATACAGTGCAGTAAAGCTCTGCTTGTGAGACTTC - 27360
 - L L D A T Y T L T V K S L A L S D E D F
 - C S L P R I H * R * S P L L * A N K T S
 - A P C H V Y T D G K V P C F E R * R L H
 27361 - ACCTRAGTTGAGTGTGCACTTTGCGCAGCGATAGTGACTTGATCAATGCACRTTC - 27420
 - T * V E * S Q L C A S D S D L I N A H F
 - P K L S D R N F A P A R I V T * S M H I S
 - L S * V I A T L R D Q * * L D Q C T F R
 27421 - GAGTCCTCTTAAACACATCAATGAAGCATTTAACACAACTTGTGTTATGAGC - 27480
 - E C L V N N N I N E A F Y T I L D V I * S
 - S A L L T T S M K H F T Q S L N L S E A
 - V F C * Q H Q * S I L H N P * C Y L K Q
 27481 - AACCTGTTATGACCTCTGACCTGACATGTCACAAAACACCTGTAAATGAGAAATTGAGAATCTC - 27540
 - N L Y L T L D D V K N T C N E K F E N L
 - T C I * P L T M S K T P V M R N L R I S
 - P V F E D P * R C Q K H L * * E I * E S P
 27541 - CCAAGCATCTTGAGAAATTCAACCTCCGTGACTTGTTGCGCTCAATCCATTAAAGAT - 27600
 - P S I L K F N S C T K F R L N P F K D
 - Q A S L R N S T P A L S F A S I H S K I
 - K H P * E I Q L L H * V S P Q S I Q R *
 27601 - AGGCTGAGTTTCACAGTAGTGTGCCAAAAGATAAGACACRACCTGAGAAGTGTG - 27660
 - R P E F F N S S A Q K I R Q P L R S L L
 - G L S F S T V P K R L D N H * E V C C
 - A * V F Q Q * C P K D * T T T E K S V V
 27661 - TACAAAGACCAACAGGTACATAGGCCATAATAGACACTGTTGGTGGAGCAGGTCTGAGT - 27720
 - Y K T P V T Y I C H N N D T V G E Q V * S
 - T R P F V T Y I A R I M T L L V S R S E V
 - Q D H Q L H M P * * H C W * A G L K Y

FIG. 12 Con't

88/106

27721 - ATAAACCATTGGCGTCGACAAGACGTAATGACTGTTAGAATACCATCRAAGTATGGTGAC - 27780
 - I N H G V D K T * + L F R N T I K Y G D
 - * T M A S T R R N D C S E I P S S M V T
 - K P W R Q R Q D V M T V Q K Y H Q V W * Q
 27781 - AGCTGCTCTTGCRAATCAGGAATTGAGTGGTTGCTGCATCRAGTGTCGCGCRAARAT - 27840
 - S C S L Q I R N * V V C C I K C A R K N
 - A A L C K S G I E W F A A S S V R A K I
 - L L F A N Q E L S G L L H Q V C A Q K L
 27841 - TGATCTGATAACCAACGAGCGCTGTGAGGGAAAACACACAGTGGTTAAACTGATGCT - 27900
 - * S D N T S S L * G K T T Q W C * N * S
 - D L I T P A A C E G K P H S G V K T D L
 - I * * H Q Q P V R E N H T V V V L K L I S
 27901 - CTGTTGTCRATGTTCAAGCACCTTACGGGCTTCCCTGGTAACCTTATAGTTAC - 27960
 - L L S N V P S T F Y G L S L G N F I V T
 - C C P M F Q A P F T G F P L V T L * L P
 - V V Q C S K H L L R A F P W * L Y S Y R
 27961 - GCAGGCTCAACATGGTTGAAAGACTTGTAAAGACTCTTATGTCAGACTCTTATAGTTAC - 28020
 - A G L N N G F E R L V I K T L Y S V N K
 - Q D S T M V L K D L * S R L F I V S I K
 - R T Q Q W F * K T C N Q D S L * C Q * R
 28021 - GGCACATGTTAGAACGAGAACAGATGCCAAATGTCAGGCCAACCTCTCATTCATCAATGAAA - 28080
 - G T C R S R E C Q N D G N L F I Q M K
 - A L V E A E K D A K M M A T S S F E * K
 - H L * K Q R E M P K * W Q P L H S N E N
 28081 - ATCGCCACATGTTATGTTAACAGCTCACGACTCAGATCTCAAGGAGATCTCATT - 28140
 - I A N N V N V N T F T T Q Y L K E I L I
 - S P T M L M L T R S R L S I S R R S S F
 - R Q C * C * H V H D O S V S Q G D P H S
 28141 - CAAGGTCACATGTTACCACTGACAGTGGCCAGTATGGCTGAGCCAATATCAGCAGTACG - 28200
 - Q G L H I V T S N A M S A * A N I S T S
 - K V S T L S P V M P V W P E P I S A L A
 - R S P H C H Q * C Q Y G L S Q Y Q H * H
 28201 - ACAGGGAACCCAGTAGGACCTTATTATAGCAGGAAACATAGGCAACACACAGCCCTC - 28260
 - T R N P V G T L I I R A N I G K H T A S
 - R G T Q * A R L L * Q P T * A N T Q P P
 - E E R H A Y S S Q H R Q T H S L Q
 28261 - AAAACATCTAGTCTACCTCCCTGGGAGTGGTTCAATGTTGGATGTTGGT - 28320
 - K T S S P T S L A E S S F N V * V V V I
 - K H L V L F P L R S R V S M F E W L * *
 - N I * S Y L P C G V E F Q C L S G C D N
 28321 - ATCTGCRACRACTATGTCAGGTCACATCTCTGGCTTGTACAGGCAAGACATGGCATTT - 28380
 - I C N T M L R S N L W V L T G R T W H F
 - S A T L C S G P I S G S * Q A G H G I F
 - L Q H Y A Q V Q S L G L D R Q D M A F S
 28381 - CACTACAGCATAGTAGGTAGGGTACCCCATGTTAGGTAGGGTCTTCATRACTAATTTTC - 28440
 - H Y S I S R * V P T C S R S F N N * I F
 - T T A L V G R Y P H V V G P S I T K F S
 - L Q H * * V G T H M * * V L Q * L N F Q
 28441 - AGTGCCACATGTTACCAAGTGGCTTCAGAAGTTCAGCAGTCGCACGCTGCCRATGAACTTCATC - 28500
 - S A T M F T S G F Q K V A R L P * N F I
 - V P Q C S Q V A F R K S H V C H E T S S
 - C H N V H K W L S E S R T S A M K L H R
 28501 - GCAAATGTTACATTCTCATCAAGTGTACAGACAGTCGATATGTTACACTCTGTGGAGATGC - 28560
 - A M I T F H Q G R Q V H I V T L L W R C
 - Q * L H F I K V W D K C I L L H S C G D A
 - N D Y I S S R * T S R Y C Y T P V E M Q

FIG. 12 Con't

89/106

- 28561 - AACAGGGTACACAGAGCGTATAGCCCCATGAAACCCCTAGCTTTTCTTTCAACACG - 28620
 - N R V H R A Y T P H E T L S L F L F N T
 - T G Y T E R I R P M K P S V F F F S T R
 - T Q G T Q S V Y A P * N H P Q S F S F Q H V
 28621 - TGGITGAATGACTTTGACTTTGAGTTAAGAGGAACACAAACTTGGCATTCCTT - 28680
 - W L N D F D F * V K R K H K L W A F P F
 - G * M T L T F E L R G N T N F G H S P L
 - V E * L * L L S * E E T Q T L G I P L *
 28681 - GAANGTCAAAATTCTGGCACCTCTTAATTGAAAGGGTGTCTGGTGTCTGAGCTT - 28740
 - E S V S N K F L G T L N F E G C L V L V A L
 - K V S N F L A L L I S K G V W C S * L L
 - K C Q I S W H S * F R R V S G A R S S Y
 28741 - ATCAGACCCCTAGTCAACAGGAATTTCATGTCATGTCAGGCACAGTGACACC - 28800
 - I R A L N F M L M V T A A V D T
 - S E R S V N Q A I S C S W S R Q Q * T P
 - Q S A Q * T R Q F H A H G H G S S R H L
 28801 - TCTCTTGACTGATGTAATCAAGTGTGCGGAAAGTGACTGACATTGCGCG - 28860
 - S L R L D V I K L F G K S A H * L A R A
 - L F D S M * S S C S E R V H I D L P A R
 - S S T C R N Q V V R K E C T L T C P R V
 28861 - TCGGAGAAAATCTTGTCAATCAAGGGTACCCATCTGGCCACAGAAATGTTGTC - 28920
 - C E K I F D A I K R V P I W A T E I V V
 - A R K S L M Q S R G Y P S G P Q K L L S
 - R E N L * C N Q E G T H T L G H R N C C R
 28921 - GACATAGCGAGTGTACTGCACCTTCATTGAGCTCACGGAGTGACACT - 28980
 - D I A T C D S T S I E L T S E F T E C T T
 - T * R V T A P P L S S R V S S R S A P L
 - H S E * L H L H * A H E * V H G V R H C
 28981 - GCGATGCTAGTGTCTTCAATGTCATGGGATCAGTCGAAGCTGTC - 29040
 - A M L S V P V L F I I F N G I S A K L V
 - P C L V F Q F C S * S S M G S V P S S S
 - H A * C S S F V H N L Q W D Q C Q A R H
 29041 - ACCTTGTCTAGTGTACTGGATCTGATGCGATGCTATGACCCCGCTCCCTATTAC - 29100
 - T * V I R L * I D A I A M T T G S L I T
 - P K S * D F R S M P * L * P P A P L L P
 - L S H K T L D R C H S Y D H R L P Y Y R
 29101 - GTTCTTACAGAGAAAGAACATGGCTATGCAATTGGGGTTGGCCACATGTCAG - 29160
 - V L T K K N I A V C N W G F A H M W H E
 - F L R R R T L R Y A I G V S P T C G T S
 - S Y E E E H C G M Q L G F R P H V A R V
 29161 - TACTCCCAGTGTATACCGCTACGCCACTGATGCGCTCATCTGCAACAGCTC - 29220
 - Y S C Y T A T T V L N A V H F C N Q L
 - T P S V I P L R P Y * M P S I S A T S S
 - L P V P L V R Y D R T E C R P F L Q P A Q
 29221 - AACGACCTTGTGGCGTGTGGCTTAAGGCATCAGAACCTTAAATGAACACATAGGG - 29280
 - N D L V A V I G A * G I R T F N E H I G
 - T T L W P * L V L K A S E R L M N T * G
 - R P C G R D W C L R H Q N V * + T H R A
 29281 - CTGTTCAAGCTGGCGTACGCCCTTTTCAGCTACTAGACCCACAATGCGCATTTT - 29340
 - L F K L G Q Y A F F Q L Y * T T S A I F
 - C S S W G S T P F P S S S T R P Q V P F L
 - V Q A G A V R L F F P A L L L D H K C H F *
 29341 - GAGGTGTTCACTGGCTCCGATAGGGCTCTTCCACAGATCCCGAAGGCCACGACTAG - 29400
 - E V F T C L R * G L F H R V P E A T H *
 - R C S R A S D R A S S T E S P K P R T S
 - G V H V P P I G P L P Q S P R S H A L A

FIG. 12 Con't

90/106

29401 - CAGGTCTCTAACCTGAAGGACAGGCAAACCTGAGTTGGACGTGTGTTTCGTTGACACC - 29460
 - H V S N L K D R Q T E L D V C F L V D T
 - T S L T * R T G K L S W T C V F S L T P
 - R L * P E G Q A N * V G R V F S R * H Q
 29461 - AAGAACAAAGGCCTCCATCTACCTTTCGGTCACACCGGACGAAACCTAGGTATGCTGA - 29520
 - K N K A L H L T F R S H P D E T * V C *
 - R T R L S I L P F G H T R T K P R Y A D
 - E Q G S P S Y L S V T P G R N L G M L M
 29521 - TGATCGACTGCAACAGGGCAGAACCGTAAGCACTCTGAGAAGAGGGCAGAGTTACTCG - 29580
 - * S T A T R T K P * A V C R R G T S Y S
 - D R L Q H G R N R K Q S A E E G R V T R
 - I D C N T D E T V S S L Q K R D E L L V
 29581 - TTCTTGTCARCGACGATRAAAAATTATATGTTTACTGCGTAGGTGCACTAGGCATG - 29640
 - F L V N D S K I Y Y C L Y C V G A L G M
 - F L S T T V K F I I V Y T A * V H * A C
 - S C O R Q * N L L F I L R R C T R H A
 29641 - CAGCCGAGCGACGCTACACAGATTAAAGTTCGGTTAGAGAACAGATCTACAGAGAT - 29700
 - Q P S D S Y T D F K V R L E N R S T R D
 - S R A T A T Q I L K F V * R T D L Q E I
 - A E R Q L H R F * S S F R E Q I Y K R S
 29701 - CGAGGTTGGTGGCTTTCTCGTAGCTAAACCTAATAT - 29742
 - R G W L A F P G * V K T * Y X
 - E V G W L F L G R * K P N X
 - R L V G F S W V G K N L I X

FIG. 12 Con't

91 / 106

N-gene primers (nucleotide position 29247-29410)

200#(5'-cttgttgtggcatcatgtgg-3')

S-gene primers (nucleotide position 24751-25049)

131#(5'-cacagaggaaacctttttt-3')

132# (5'-tcccaattttgaaggtaatgg-3')

FIG. 13

92 / 106

ATTCCTGATAATGGACCCCATAAACCAAGTACTGCCCGATTAATTCATTGGACCCCCAGATTCACATGCACTGACAAATGCCAGAATGGAGGACG
 CAATGGGGCAAGGCCAAAAAGCCGACCCCAAGTTACCCAATAATACCTCGCTCTTGTCAGCCTGCTCACTCGATGCCAAAGGAACTTA
 GATTCCTTCAGGCCAGGGGTTCAGATCAACCAAATAGTGGTCAAGTGGCAAAATGCTGACTAACCGAGACTGTTGGCTGTGGTGT
 GAGGGCAAAATGAAAGGCTAGGCCAGAGCTTACCCAGATGGTACTCTTAACTTACCCAGGAGCTTACCCAGGAGCTTACCCAGGAGCTT
 CGATGGGTGCACTGAGGGAGCTTAAATACCCAGAGCCAATTCAGGACCCCAATCTTAATACAAATGTCACCCGCTGCTCAACTCTCT
 TAAAGGAAACAAATGCCAAGGGCTTACGCCAGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 AAATCAACTCTGCCAGCACTAGGGAAATTTCCTGCTCGATGGCTAGGGAGGTGGAAACTGGCTGGCTGAGATGCTGAGCTGATGAA
 CCAGCTTGAGAGCAAACTTCTGTCAGAATTTGCTGCTGAGGCTAAAGGCTGCTAAACCTGCGCAAAAGCTACTGCGCACAAACGTACAACTCA
 CARACAAAGGCCAAACTGTCAGAATAGAATTTGCTGCTGAGGCTAAAGGCTGCTAAACCTGCGCAAAAGCTACTGCGCACAAACGTACAACTCA
 AGGATTTGGGAGAGTGTGTCAGAACAAACCCAAAGGAAATTGGGAGCAAAACCTTACGAGCAAGGAATCTGATTAACACTGGCGCAATIG
 CAAAAATTCCTCAAGGGCCCTGCACTCTGGATGTCACCCATGGATGGAAAGTCAACCTTCGGAAACAAAGGCTGACTTATCTGAGCCATT
 AAATGGATGCAAGTCAAGTCAAAAGCAAGCTTAACTCAAGGCAAGTCAATCTGAAACCAAAATTCAGCTGAAACCAAAATTCAGCTGAA
 GGCAAAAGAAAGAAAGACTGATGAGCTGGCGAGAGAACAGGAGCCACTTGACTCTGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGG
 TCCTGAGAAGTCAAACTCATGAGGGAGCTCTGCTGATTCAACTCAGGGCTATA

FIG. 14A

93 / 106

MSUDNGPOSNORSAPRPRITFGGPTDSTDNNONNGGRNGARPKORRPGQLBNNTASWFTALTQHGKEELRFPRGGVPIINTNSGPDOIGYYRATRVRGG
DGKMKELSPRYYEYLIGTGEASLIPYGAKEGIVWVATEGALATPDKHIGTRNPNNNATVQLLPQGTTLPKGYAEGSRGGQASSRSSRSRGNSR
NSTPGSSRGNSPARVASGGGETALLLIDRNQLESKVSKGKQ00QGQTVTKSAAEASKPRAKRTATKOVNTPQAFCRRGPECTQGNFGDDLIR
QGTDYKHWQPAQFAPSASAFFGMSNLIGEVIPSGLWLTZHGATKLQDKDQFQKDNTVILINKHIDAYKEPTEPKDKKKTDEAQPLPQRQKQPT
VLLPAAMDFSTQLQSMSASASTDA

FIG. 14B

94 / 106

FIG. 15A

95 / 106

FIG. 15A Con't

96 / 106

MFTFLIIFIITSGSGLDRCTTEDVQAPNYQHTSSMRGVYYPDEIERSDTLYLTDLFPLPEYSNTVGETTINHTFGNPVTPEKDGIYPAATEKSNNVVRGVFESTANNKOSVUINNNTNVIRACNFELCDNPEFAVKPMGQTTHMIFDNAFNCTFEYISDAFSLDVSEKSGNFKHREFVFKNDKGFLXVYKQPIDVVRDLPSEGNTLKPIKPLGININTNPRALITAESPQDGNTSAALAEVGTLKEPTFMKVDENGTTDAVDCSQNPLABIKCSVKSPEIDEGITQTSNFRVVPSPGDVURFPNTNLCPEGEVENATREPSTYAWERKISNCVADSYLYNSTFFTCYGSAPKLNLCESNYADSFVWGDIDVQIAPIRGQTGVIAIDNKKLEPDDMMGCVLANTRNLIDATSTGNYNKXYLRLKGKLPFERDISNVPSPGDKEPTPPALNCWPANDYGFXTTGTGIVQYQRUVVLSFELINAPATVCGPKLSTDILRNQCNENFNGLTGVLIPPESSRKFQFOFGRDUVSDETDVSDRDPKTSEILOISPCSFGGVSVITPSTNASESEAVLYQDVNCIDVSTATHADQLTPAWRIYSTGNVFQTOAGCCLIGAEHVDSYECUDIPIGAGCASYHVSQSLRSQSIVATMSLGADSSIASNNNTIAPTNFSISITBPUVPMVSYMAKTSVDCNMVTCGDBTBANLILONGSECTONRALSGTAEQDRNTRYFAQVKOMYKPTLKYEQQGENESQILPDPLKPTKRSFEDILLENKVTLADGFEMKOYGEGLGDINARDLICADEKENGLTVPPLTDMIAITYALVSTATAITFGAGAALQIPFAQMOMATRENGIGTQVNLYENOKQFNAKAIISQIQCSEITMTSIALKGKQDQVNQNAQALNTLVQULSSNFGAISVVLNDIISRLDKVEAEVQIDRLITGRLOSLQTYVQQLRAEIRASANIAKTMSECVIGQSKRVDIFGKEYHIMSFPOAAFHGVVFLAVTVPSQERNFTTAPACHEGKAYPREGVEVFGNGISWEITQNEEFPQUITTNTVEVSGNCDVYVIGINNTVYDPIQPELDSTKEELDKBFKNHSTSPPDVUDISGINASVVMNIOKEIDRNEVARLNESLIDQELGKYEQVTKNEWTVWLGFLAGLIAIVMVTTIICCMTCSCSCLKGACSGSCCKFDEDDBSEPVLKGVKLYT

FIG. 15B

97/106

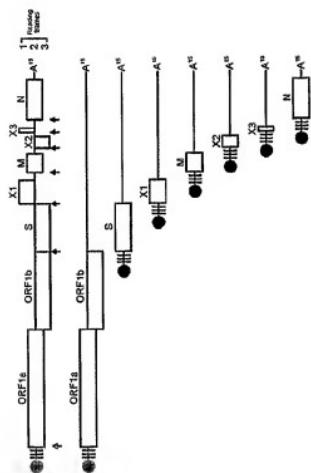


FIG. 16

98/106

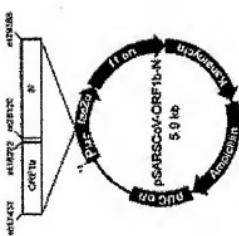


FIG. 17

99/106



FIG. 18

100/106



FIG. 19

101 / 106

- Primers for a 225-bp fragment of the region of N-gene that showed no homology to other coronavirus:

SRS251: 5'-GCAAGTCAAAGCCTCTTCG-3' (SEQ ID NO:2480)

SRS252: 5'-GCCTCAGCAGAATTC-3' (SEQ ID NO:2481)

- Primers for 181-bp fragment of the region of 1b-gene:

coro3: 5'-TACACACCTCAGCGTTG-3' (SEQ ID NO:3)

coro4: 5'-CACGAACGTGACGAAT-3' (SEQ ID NO:4)

- Primers for a 745-bp fragment from pig β-actin gene:

Actin-F: 5'-TGAGACCTTCAACACGCC-3' (SEQ ID NO:2482)

Actin-R: 5'-ATCTGCTGGAAAGTGGAC-3' (SEQ ID NO:2483)

FIG. 20

102/106

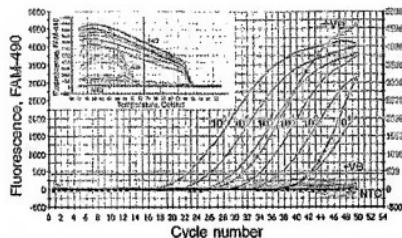


FIG. 21A

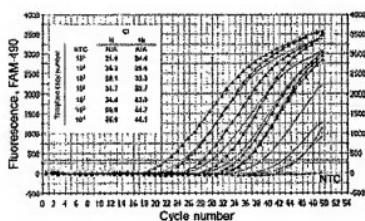


FIG. 21B

103/106

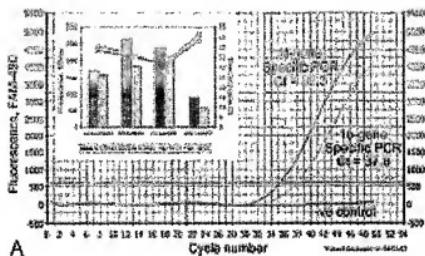


FIG. 22A

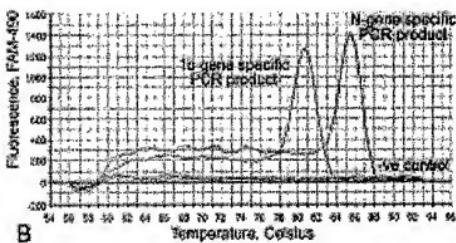


FIG. 22B

104/106

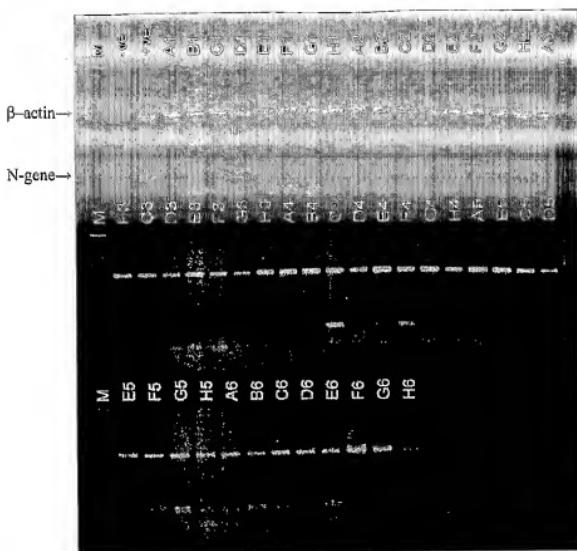


FIG. 23

105/106

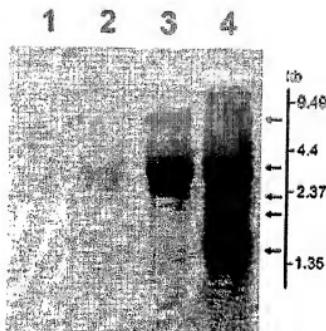


FIG. 24

106/106

FIG. 25

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2004/000246

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N 15/50, C07K 14/165, A61K 39/215, G01N 33/569, A61P31/14, A61P11/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N A61K C07K G01N A61P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, WPI, EPODOC, PAJ, GENBANK, PUBMED(SARS, CORONAVI+ etc.)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JOURNAL OF VIROLOGY, Vol. 74, No. 19, Oct. 2000, P 8913-8921 NATHANIE ARBOUR ET AL.: Neuroinvasion by Human Respiratory coronaviruses" see the whole document	1-19,25-29

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubt on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
09 JUN. 2004(09.06.2004)

Date of mailing of the international search report

01 · JUL 2004 (01 · 07 · 2004)

Name and mailing address of the ISA/CN
6 Xitucheng Road, Jimen Bridge, Haidian District,
Beijing, 100088, P.R. China
Facsimile No. (86-10)-62019451

Authorized officer

DING, Huiping



Telephone No. (86-10)-62085300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2004/000246

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CLINICAL AND DIAGNOSTIC LABORATORY IMMUNOLOGY, Vol. 9, No. 6, Nov. 2002, p 1392-1395 ARLENE R. COLLINS. : "In Vitro Detection of Apoptosis in Monocytes/Macrophages Infected with Human Coronavirus" see the whole document	1-19,25-29
PX	CN1482461A 17 MAR. 2004(17.03.2004) see claims see examples	2,4,4,5,7,9-11, 27-29
PX	CN1449826A 22 OCT. 2003(22.10.2003) see claims see examples	2,4,5,7,9-11
PX	CN1465589A 07 JAN. 2004(07. 01. 2004) see claims see examples	2,4,5,7,9-11
PX	CN1472332A 04 FEB. 2004(04. 02. 2004) see claims see page 3, line 24-page 4, line 2	1,4,5,7

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/CN2004/000246**Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item itemI.b of the first sheet)**

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:
 - a. type of material
 - a sequence listing
 - table(s) related to the sequence listing
 - b. format of material
 - in written format
 - in computer readable form
 - c. time of filing/furnishing
 - contained in the international application as filed
 - filed together with the international application in computer readable form
 - furnished subsequently to this Authority for the purposes of search
2. In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORTInternational application No.
PCT/CN2004/000246**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:20-24
because they relate to subject matter not required to be searched by this Authority, namely:
Claims 20-24 are directed to diagnostic methods.
2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on protest The additional search fees were accompanied by the applicant's protest. No protest accompanied the payment of additional search fees.